Genome-wide dysregulation of histone acetylation in the Parkinson's disease brain

Lilah Toker, PhD[1,2†], Gia T Tran, MD[1,2†], Janani Sundaresan, MSc[1,2], Ole-Bjørn Tysnes, MD, PhD[1,2], Guido Alves, MD, PhD[3,4], Kristoffer Haugarvoll, MD, PhD[1,2], Gonzalo S Nido, PhD[1,2], Christian Dölle, PhD[1, 2] and Charalampos Tzoulis, MD, PhD[1,2*]

1. Neuro-SysMed, Department of Neurology, Haukeland University Hospital, 5021 Bergen, Norway. 2. Department of Clinical Medicine, University of Bergen, Pb 7804, 5020 Bergen, Norway. 3. The Norwegian Centre for Movement Disorders and Department of Neurology, Stavanger University Hospital, Pb 8100, 4068 Stavanger, Norway. 4. Department of Mathematics and Natural Sciences, University of Stavanger, 4062 Stavanger, Norway.

[†] These authors contributed equally to the work

[*]**Correspondence to:**
Prof. Charalampos Tzoulis
Neuro-SysMed Center of Excellence for Clinical Research in Neurological Diseases
Department of Neurology, Haukeland University Hospital
Department of Clinical Medicine, University of Bergen
5021 Bergen, Norway.
Telephone: +47 55975061
E-mail: charalampos.tzoulis@nevro.uib.no
E-mail-2: charalampos.tzoulis@helse-bergen.no

Number of words in body of manuscript: 5,098

Number of display items: 9

Keywords: H3K27, epigenetics, neurodegeneration, ChIP-seq, sirtuin, mitochondria, gene expression

List of abbreviations:

ChIP-seq: chromatin immunoprecipitation sequencing

Ctrl: control

DAR: differentially acetylated region

EGL: effective gene length

GWAS: genome-wide association study

HAT: histone acetyltransferase

HDAC: histone deacetylase

L-DOPA: levodopa

mtDNA: mitochondrial DNA

MSPs: Marker Site Profiles

NAD: nicotinamide adenine dinucleotide

NBB: Netherlands Brain Bank

PD: Parkinson's disease

PMI: postmortem interval

PW: Park West

RiP: reads in peaks

RLE: Relative Log Expression

SNc: substantia nigra pars compacta

SIRT: sirtuin

TSS: transcription start site

WB: Western blot

WES: whole exome sequencing

Abstract

Parkinson's disease (PD) is strongly associated with mitochondrial dysfunction. Based on the established coupling between mitochondrial energy homeostasis and protein acetylation, we hypothesized that aberrant histone acetylation may be implicated in PD. To test our hypothesis, we assessed histone acetylation in brain tissue from two cohorts of PD patients and controls, using immunoblotting and chromatin immunoprecipitation sequencing (ChIP-seq). Our findings show increased acetylation at several histone sites in PD with the most prominent change observed in the prefrontal cortex for H3K27, a marker of active promoters and enhancers. We demonstrate that PD-associated H3K27 hyperacetylation is genome-wide and has a predilection for genes implicated in idiopathic and/or monogenic PD, including *SNCA*, *PARK7*, *PRKN*, and *MAPT*. Integration of the ChIP-seq with transcriptomic data revealed that the correlation between promoter H3K27 acetylation and gene expression is attenuated in the PD samples, suggesting that H3K27 acetylation may be decoupled from transcription in PD. Our findings replicate across two independent cohorts and strongly suggest that dysregulation of histone acetylation plays an important role in the pathophysiology of PD.

Introduction

Parkinson's disease (PD) is the second most common neurodegenerative disorder, affecting ~1.8 % of the population above the age of 65 years[1]. The neuropathological hallmark of PD is the progressive loss of dopaminergic neurons in the *substantia nigra pars compacta* (SNc) in the presence of α-synuclein-positive inclusions termed Lewy pathology. Additional neurodegenerative changes occur in multiple regions of the nervous system, including several brainstem nuclei, the olfactory bulb, hippocampus, amygdala and the neocortex, as well as the autonomic and enteric nervous systems[2].

The etiology of PD remains largely unknown. While monogenic forms of PD exist, they are rare, generally accounting for less than 5-10% of all cases in most populations. The vast majority of patients have "idiopathic PD", which is of complex etiology[3]. Genome-wide association studies (GWAS) have identified multiple risk *loci* associated with idiopathic PD. These make a relatively small collective contribution to the disease risk, however, and have not been linked to specific molecular mechanisms that could be targeted therapeutically[4].

PD is strongly associated with mitochondrial dysfunction. Mutations in the *PRKN*, *PINK1* and *PARK7* genes, involved in mitochondrial quality control cause autosomal recessive PD[3]. Furthermore, multiple mitochondrial defects have been shown in the PD brain, including impaired mitochondrial DNA (mtDNA) maintenance[5,6], aberrant mitochondrial quality control[7] and defective mitochondrial respiration, predominantly in the form of widespread complex I deficiency[8,9].

It has been shown that mitochondrial complex-I deficiency can result in epigenetic alterations in the form of increased histone acetylation[10–12]. This effect is believed to be primarily mediated by decreased NAD^+/NADH ratio, leading to reduced activity of the NAD^+-dependent class III histone deacetylases known as sirtuins[10,13–16]. With regard to PD,

previous studies demonstrated decreased sirtuin activity and hyperacetylation of several histone sites, including H3K9, H3K27 and H2BK15[12], following treatment of neuronal cells with PD-associated mitochondrial complex-I inhibitors[10–12]. Furthermore, reduction in NAD^+ availability, decreased sirtuin function and histone hyperacetylation have all been shown to occur during aging[16,17], which is a prerequisite for idiopathic PD.

Despite the strong link between defective mitochondrial function and altered histone acetylation, and the emerging role of epigenetics in modulating the risk and progression of complex disorders, only a few small-scale studies of histone acetylation have been reported in PD, providing inconclusive results[12,18,19]. Moreover, the genomic landscape of histone acetylation in PD has never been assessed. Here, we report the first genome-wide study of histone acetylation in the PD brain, based on two independent cohorts from the Norwegian ParkWest study (PW)[20] and the Netherlands Brain Bank (NBB). Our findings reveal that H3K27 acetylation is profoundly dysregulated at a genome-wide level in the PD brain.

Results

Increased histone acetylation and histone protein levels in the PD brain

To determine whether the global protein acetylation landscape is altered in the PD brain, we first performed Western blot analyses in the prefrontal cortex of 13 individuals with PD and 13 demographically matched controls, using an antibody against acetylated lysine (Fig. 1A, Supplementary Table S1). We concentrated on the prefrontal cortex because this region is affected by the disease, but is less likely to be biased by changes in cell composition (i.e. neuronal loss and gliosis), which confound observations in severely affected areas such as the SNc[2]. Linear regression analysis of the densitometric measurements adjusted for age, sex and PMI indicated an increased acetylation signal at ~17 kDa, consistent with the molecular weight of core histone proteins ($p = 0.014$, Fig. 1A).

To confirm that the hyperacetylated proteins were indeed histones and to dissect histone- and site-specificity, we performed an in-depth analysis of specific acetylation sites across the four core histones: H2AK5ac, H2BK15ac, H3K9/K14ac, H3K27ac, H3K56ac, H4K5ac, H4K12ac, and H4K16ac. The acetylation status of α-tubulin, a cytoskeletal protein typically regulated by acetylation, was assessed as a non-histone control. To maximize sensitivity, we focused our analyses on the 7 PD and 7 control samples showing the strongest difference in the pan-acetylation signal (Supplementary Table S1). The acetylated fraction of each histone site (the ratio between the signal densities of the acetylated histone and total histone protein) as well as total histone quantity were compared between PD and control samples using linear mixed models with gender, age and PMI as fixed effects and subject and blot number as random effects. These analyses indicated a significant increase in H2BK15ac ($p = 1.4 \times 10^{-5}$), H3K9/14ac ($p = 3.9 \times 10^{-4}$), H3K27ac ($p = 4.5 \times 10^{-8}$), H3K56ac ($p = 0.01$) and H4K12ac ($p = 5.0 \times 10^{-3}$) in PD. No significant difference was found in the acetylated fraction of H2AK5, H4K5, H4K16, or α-tubulin (Fig. 1B-C, Supplementary Table S2). Interestingly, the total

quantity of the four core histone proteins was also elevated in PD (H2A: $p = 2.7 \times 10^{-5}$, H2B: $p = 1.1 \times 10^{-8}$, H3: $p = 2.0 \times 10^{-7}$; H4: $p = 3.9 \times 10^{-18}$).

While we did not expect major changes in cellular composition in the prefrontal cortex of PD patients, we sought to rule out this potential bias in our samples. To this end, we carried out immunoblot analyses of several cell-type markers: NeuN (neurons), GFAP (astrocytes), CNP1 (oligodendrocytes) and CX3CR1 (microglia) as well as the nuclear marker NUP188, in the same samples used for acetylation assessment. The only significant change was observed for CNP1, showing increased signal intensity in PD samples ($p = 3.9 \times 10^{-5}$, Fig. 1D, Supplementary Table S2). We reasoned that subcortical white mater carry-over during dissection could result in this interindividual variation in the cellular composition of our samples. To ensure that our findings were not driven by increased oligodendrocyte abundance in the PD samples, we repeated our prior immunoblot analyses adding CNP1 as a covariate to the statistical model. As shown in Supplementary Table S2, all of the indicated hyperacetylated sites remained significant after CNP1 adjustment.

In order to determine the anatomical specificity of H3K27ac and H3K9/14ac in the PD brain, we assessed these two markers in the cerebellum and striatum from the same patients. H3K27 was significantly hyperacetylated in the PD striatum ($p = 0.034$) and cerebellum ($p = 0.016$). H3K9/14 was significantly hypoacetylated in the PD cerebellum ($p = 7.0 \times 10^{-3}$) and unchanged in the striatum ($p = 0.7$, Fig. 2, Supplementary Table S2).

Histone hyperacetylation in the PD prefrontal cortex is accompanied by upregulation of sirtuin protein levels

Our study was motivated by the hypothesis that the well documented complex-I deficiency in the PD brain is likely to decrease the NAD^+/NADH ratio, thereby reducing the activity of sirtuins, the NAD^+-dependent histone deacetylases. A direct confirmation of this hypothesis

would require a reliable determination of intracellular NAD^+ levels in our samples, which is not feasible due to its rapid degradation in postmortem tissue[21]. However, it was previously reported that decreased sirtuin activity caused by NAD^+ deficiency is accompanied by increase in the protein quantity[22,23]. Thus, we assessed whether sirtuin protein levels were altered in the samples tested for histone acetylation. We focused on SIRT1, SIRT2 and SIRT3, localized to the nucleus, cytosol and mitochondria, respectively, since they have been linked to aging and neurodegeneration[13]. In line with our hypothesis, immunoblot analysis indicated a significant upregulation of SIRT1 ($p = 2.0 \times 10^{-3}$) and SIRT3 ($p = 3.0 \times 10^{-3}$) in PD samples. No difference was detected for SIRT2 ($p = 0.57$, Fig. 2C, Supplementary Table S2).

Histone hyperacetylation is not induced by anti-Parkinson drugs *in-cellulo*

We next investigated whether the observed altered histone acetylation in PD could result from treatment with anti-Parkinson agents. To test this, we exposed fully differentiated SH-SY5Y cells to the drugs used by most individuals with PD from the PW cohort during their last year of life (L-DOPA, carbidopa, entacapone and ropinirole), and monitored acetylation of H3K9/14 and H3K27. Immunoblot analysis revealed no effect for any of the tested anti-Parkinson drugs on histone acetylation (Fig. 2D).

Acetylated H3K27 exhibits higher genomic occupancy in PD

Our immunoblot analyses identified H3K27ac as the modification showing the most pronounced alteration in the PD prefrontal cortex. To gain insight into the genomic regions undergoing this alteration, we carried out ChIP-seq analysis of H3K27ac in samples from the PW cohort (discovery cohort, Supplementary Table S1).

First, we characterized the genomic distribution of H3K27ac in the PD and control groups. For this purpose, we identified H3K27ac enriched genomic regions (peaks) in each group separately (see methods). In line with the increased H3K27ac signal observed in the immunoblot analysis, higher number of peaks (146,763 vs. 135,662), as well as a higher proportion of group-unique peaks (32.2% vs. 13.2%), and wider genome coverage (10.6% vs. 9.11%) were found in the PD compared to the control group (Fig. 3A, Supplementary Table S3). We then compared the distribution of genomic annotations and detection p-values of overlapping and unique peaks between the PD and control groups. In both groups, unique peaks occurred more frequently in intergenic and intronic regions than in promoter and exonic regions (Supplementary Fig. S1A,B). One possible interpretation for this observation is that unique peaks show a predilection for distal enhancer regions. However, the unique peaks were also characterized by significantly higher detection p-values, compared to the overlapping peaks (Fig. S1C,D). Thus, we believe that a more plausible explanation is that most unique peaks represent non-targeted H3K27 acetylation.

ChIP-seq data corroborates H3K27 hyperacetylation in the PD brain

Next, we redefined H3K27ac peaks by identifying H3K27ac enriched regions in aggregated samples (see methods). The final peak-set consisted of 160,521 peaks, covering 11.6% of the genome and distributed over all 22 autosomes and both sex chromosomes (median peak size, 817bp, Fig. 3D). All downstream analyses in the discovery cohort were performed based on this peak-set, further referred to as PW_peak-set.

ChIP-seq data reflects the abundance and genomic distribution of the target protein (or protein modification) which can vary substantially between samples[24]. Thus, the inter-individual variation in ChIP-seq library size represents a combination of noise and genuine biological variation. In our study, we anticipated an increase in overall H3K27ac ChIP-seq

signal in PD, based on the observed increase in H3K27ac fraction in the Western blot analysis. We therefore reasoned that the most suitable approach for normalization of our data would be using reads in peaks (RiP) annotated to house-keeping genes[24]. We modified this approach by manually selecting peaks annotated to known housekeeping genes and exhibiting a high and broad peak signal covering most of the gene body. The rational here was that these genes should be saturated with histone acetylation and are, therefore, expected to exhibit little biological variation. Thus, any observed variation in these peaks is likely to represent technical variation and/or noise. Supplementary Fig. S2 shows the H3K27ac ChIP-seq peaks annotated to the previously proposed[24] and the chosen housekeeping genes.

To validate our normalization approach, we performed additional H3K27ac immunoblot and ChIP-seq analyses in samples from 55 individuals (including the subjects from our discovery cohort, Supplementary Table S1) and tested the correlation between the immunoblot and ChIP-seq data normalized by different approaches. As shown in Supplementary Fig. S3, the highest correlation between immunoblotting and ChIP-seq data was observed when the total number of RiP was normalized to our manually chosen peaks. Importantly, comparison of normalized RiP between PD and control samples corroborated the finding of general H3K27 hyperacetylation in PD (Supplementary Fig. S3).

Cell composition is a major source of variation in bulk tissue H3K27ac ChIP-seq data

It is established that different cell-types exhibit distinct epigenetic landscapes[25]. Therefore, bulk tissue data must be evaluated in the context of the underlying cellular composition[26–28]. In order to estimate the relative cellular composition in our ChIP-seq data, we used an approach similar to the one described by Mancarci et al.[27]. Briefly, we intersected genomic H3K27ac binding sites differentiating between NeuN-positive and NeuN-negative brain cells[25] with the list of brain cell-type specific marker-genes[27] to obtain genomic H3K27ac

cell-type marker sites. The first principal component of reads aligned to the marker sites was used to obtain Marker Site Profiles (MSPs) which serve as a proxy for the relative abundance of the corresponding cell-types across the samples. To validate this approach, we analyzed publicly available ChIP-seq data from cortical and cerebellar bulk tissue samples[29], as well as ChIP-seq data from the entorhinal cortex of individuals with Alzheimer's disease and healthy controls[30]. MSP analysis successfully recapitulated the well-known increased abundance of glial cells in cortical vs cerebellar samples[31,32], and decreased number of neurons in the entorhinal cortex of subjects with Alzheimer's disease compared to controls[33,34] (Supplementary Fig. S4A,B).

Next, we performed MSP analysis in samples from the PW cohort. This analysis indicated no significant differences in cell-type abundance between the PD and control groups (Supplementary Fig. S5). Nevertheless, based on our findings of variable oligodendrocyte content in the subset of subjects tested by immunoblotting, we chose to assess the effect of oligodendrocyte MSPs, as well as other experimental and demographic covariates, on our ChIP-seq data. Strikingly, sample-to-sample correlation of the PW_peak-set data showed that the samples clustered mainly based on oligodendrocytes MSPs (Fig. 3B). Accordingly, the first principal component of the data was mostly associated with oligodendrocyte MSP (Supplementary Fig. S6). These results confirmed that that oligodendrocyte composition is a major source of variance in our data. Therefore, oligodendrocyte MSPs were accounted for in all downstream analyses.

Hyperacetylation of H3K27 in the PD brain is a genome-wide phenomenon

We next sought to identify differentially acetylated regions (DARs) for H3K27 between PD and control samples in our discovery cohort. To further constrain our analysis to high confidence regions, we only included peaks with an average coverage > 5 counts/base (see

methods). After filtering, 133,716 peaks annotated to 17,182 genes remained for downstream analyses. Five samples (one control and four PD samples) were detected as outliers and excluded from differential peak analysis (Supplementary Table S1). Differential peak analysis was performed using the "DESeq2" R package, adjusting for age, sex, batch, PMI, oligodendrocyte MSP and normalization peak ratio. DARs were defined as differentially acetylated peaks with adjusted p-value < 0.05.

Our analysis identified 2,877 hyperacetylated and only 14 hypoacetylated PD-associated DARs, annotated to 1,434 and 9 genes, respectively. An MA plot (the log ratio vs. mean count) of all regions included in the analysis indicated a general trend for H3K27 hyperacetylation in PD, independent of the number of reads aligned to a region (Fig. 3A), strongly suggesting a genome-wide hyperacetylation of H3K27. The top-ranked hyperacetylated and hypoacetylated DARs resided within the *DLG2* (adjusted $p = 9.8 \times 10^{-4}$) and *PTPRH* (adjusted p-value 9.0×10^{-3}) genes, respectively, both of which have been previously associated with PD[4,35–37]. The full list of DARs with their annotated genes is provided in Supplementary Table S4.

H3K27 hyperacetylation in the PD cortex is replicated in an independent cohort

In order to validate our findings from the discovery cohort, we carried out ChIP-seq analysis in an independent cohort of subjects (11 controls, 10 PD) from the Netherlands Brain Bank (NBB, Supplementary Table S1). Compared to the discovery cohort, we identified overall less H3K27ac peaks and a slightly smaller genomic coverage in NBB (109,183 peaks after filtering, annotated to 17,306 genes, Supplementary Table S3). Nevertheless, when peaks were called for each group separately, we observed an increase in the percentage of unique peaks as well as increased total genomic coverage in the PD group (Supplementary table S3), confirming the trend for H3K27 hyperacetylation observed in our discovery cohort. The

distribution of genomic annotations among the unique and overlapping peaks and the distribution of peak p-values were similar between the two cohorts (Supplementary Fig. S1). Similar to the discovery cohort, cellular composition accounted for most of the interindividual variance in this cohort (Supplementary Fig. S6B, S7A).

Differential acetylation analysis replicated the finding of PD-associated genome-wide H3K27 hyperacetylation, as seen in the discovery cohort (Supplementary Fig. S5B). In total, the analysis identified 2,486 hyperacetylated DARs, mapped to 946 genes, and 227 hypoacetylated DARs, mapped to 253 genes (Supplementary table S5). Strikingly, 275 out of the 946 genes with hyperacetylated DARs and 2 out of 253 genes with hypoacetylated DARs replicated across both cohorts (Fig 4C, hypergeometric test: $p = 2.29 \times 10^{-85}$ and $p = 7.9 \times 10^{-3}$, respectively). These 277 genes are henceforth referred to as "replicated genes". The top replicated genes (ranked based on their adjusted p-value in the PW cohort) are shown in Table 1 and a complete list is provided in Supplementary Table S6.

Meta-analysis of ChIP-seq data from PW and NBB cohorts

In order to perform meta-analysis of the ChIP-seq data we first harmonized the data from the PW and NBB cohorts by pairing peaks with the same structural and functional genomic annotation (e.g. promoter, exon, 5'UTR) in both cohorts. These peaks are henceforth referred to as "common peaks". For example, two peaks were considered common if both were annotated to the first exon of the *PTPRH* gene, but not if one was annotated to the first exon of *PTPRH* while the other to the second exon of this gene (Fig 4A,B). For each common peak pair, we then calculated the Fisher's meta p-value (metaP).

In total, the harmonization process resulted in 52,286 pairs of common peaks. Out of these, 7,288 (mapped to 2,972 genes) had adjusted metaP < 0.05 (Supplementary Table S7, Fig. 4D,F). Moreover, 2,837 and 2,647 pairs included peaks defined as DARs in the PW and

NBB cohorts, respectively, with 300 pairs of peaks defined as DARs in both cohorts (Fig 4D). Similar to the cohort comparison at the gene level, this overlap was highly significant (hypergeometric p-value = 1.16×10^{-34}). The 300 common peak pairs, mapped to 197 genes, are henceforth referred to as "replicated DARs". Among the 300 replicated DARs, 296 exhibited the same direction of change in both cohorts: 295 were hyperacetylated and only a single DAR, mapped to the *PTPRH* gene, was hypoacetylated. Genes annotated to the top replicated DARs, ranked based on their adjusted metaP, are provided in Table 2.

Altered H3K27 acetylation has a predilection for genes with an established link to PD

We next wondered whether genes with an established involvement in PD, henceforth referred to as "PD-implicated genes", are over-represented among the genes targeted by altered H3K27 acetylation. To test this hypothesis, we compiled a list of genes that fulfilled at least one of the following criteria: (a) were associated with PD based on the most recent and largest GWAS metaanalysis[4], (b) are implicated in monogenic PD[38,39], or (c) encode proteins with a central role in PD-related neuropathology[2,40,41]. Among the 92 genes meeting these criteria (Supplementary table S8), 83 could be assessed in ChIP-seq data from both cohorts. Out of these, six genes: *DLG2*, *MAP4K4*, *CRHR1*, *MBNL2*, *SH3GL2* and *APP*, harbored DARs in both cohorts (hypergeometric p-value = 2.4×10^{-3}, Fig. 4E). In addition, 24 out of the 83 PD-implicated genes harboring common peaks had adjusted metaP < 0.05 (hypergeometric p-value = 6.1×10^{-3}, Supplementary Table S8, Fig. 4F). Strikingly, these included genes with a key-role in the pathogenesis of idiopathic and/or monogenic PD: *SNCA*, *MAPT*, *APP*, *PRKN* and *PARK7*. In addition to the pre-compiled list of PD-associated genes, the meta-analysis identified significantly altered H3K27 acetylation in genes linked to more complex forms of parkinsonism including *FBXO7* and *POLG*[38].

Differential H3K27 acetylation in the PD brain is associated with effective gene length

We noticed that many of the DARs were mapped to genes that are among the longest in the genome. We reasoned that this may be because longer genes have higher histone binding capacity and therefore are more likely to show a differential signal in the event of a generalized H3K27 hyperacetylation. To test this hypothesis, we assessed whether altered H3K27 acetylation, defined as -log10(metaP), is associated with the effective gene length (EGL), defined as the sum of all regions annotated to a gene (see methods). As shown in Figure 5A, we observed a positive association between EGL and the -log10(metaP) value for hyperacetylated regions, and a negative association for hypoacetylated regions.

Interestingly, we noted that PD-implicated genes are also characterized by longer EGL. Thus, we wondered whether the observed over-representation of PD-implicated genes among the DARs was confounded by their EGL. To this end, we assessed the association between the region's -log10(metaP) and being annotated to a PD-implicated gene, while using a linear model to adjust for EGL. The total number of common regions in a gene was included as a covariate in the model (see methods section). For hyperacetylated regions, the analysis indicated a significant positive association between altered acetylation and being a PD-implicated gene ($\beta = 0.16$, $p = 1 \times 10^{-7}$) as well as EGL ($\beta = 0.42$, $p < 2 \times 10^{-26}$). For the hypoacetylated regions, the analysis revealed a significant negative association with both being a PD-implicated gene ($\beta = -0.61$, $p < 2 \times 10^{-26}$) and EGL ($\beta = -0.24$, $p = p < 2 \times 10^{-26}$). Analysis of each cohort independently corroborated these findings, although the negative association with being a PD-replicated gene for the hypoacetylated regions was only significant in the NBB cohort. (Figure 5B, Supplementary Table S9). These results indicate that, while differential H3K27ac hyperacetylation exhibits a strong predilection for longer genes in the PD brain, this cannot fully explain the overrepresentation of PD-associated genes among the hyperacetylated DARs.

Data normalization approach and adjustment for cell-type composition greatly impact H3K27ac ChIP-seq analysis.

We next evaluated the impact of data normalization and cell-type correction on the output of our ChIP-seq analysis. For this purpose, we repeated the analysis with data using the Relative Log Expression (RLE) normalization method, implemented in the DESeq2 package, with or without MSP adjustment. As expected, when data was normalized using the RLE method, which ensures that the median fold of change of all signals in a sample relative to all samples is zero, the calculated fold change was centered around zero (Fig. 6C,D). Interestingly, based on this normalization, most DARs exhibited downregulation in PD.

Correcting for oligodendrocyte MSP dramatically increased the number of detected DARs in both cohorts, without having a substantial effect on the direction of change. This likely reflects the fact that correcting for cell-type composition decreased sample-to-sample variability due to random differences in white matter content (Fig. 3B**,** Supplementary Fig. S6, S7A). Of notice, several genes robustly exhibited DARs in both cohorts, regardless of the normalization method (Supplementary Table S10). Figure 6 shows examples for two of such genes, *PTPRH* and *DLG2*, both of which have previously been associated with PD based on genetic and functional studies[4,35–37] and were found in the top list of replicated DARs (Table 1,2).

Promoter H3K27 acetylation is decoupled from gene expression in PD

While it is well accepted that H3K27ac is a marker of active promoters and enhancers[42] and that modifications of this mark are predictive of gene expression levels[43], it was previously shown that widespread histone hyperacetylation induced by potent HDAC inhibition does not lead to major changes in gene expression, possibly due to compensatory feedback

regulation[44,45]. To investigate whether a similar decoupling occurs in the PD brain, we analyzed RNA sequencing (RNA-seq) data from the prefrontal cortex of the same individuals included in the ChIP-seq analysis (Supplementary Table 1, Supplementary Methods). For each gene, we calculated the Pearson's correlation between H3K27ac state (ChIP-seq) and transcript levels (RNA-seq) across controls or individuals with PD. We concentrated on peaks annotated to promoter regions, since these are expected to highly correlate with gene expression. For both ChIP-seq and RNA-seq data the correlation was performed between covariate-corrected signals. The covariate corrections were based on the models used for differential analysis and included age, sex, PMI, batch and cellular composition (MSPs).

As expected, the distribution of correlations between promoter H3K27ac and gene expression in the control group was shifted towards positive values (Fig. 7). In the PD group, however, the distribution of correlations was centered significantly closer to 0, regardless of the absolute correlation threshold chosen (Fig. 7). Namely, a higher threshold for absolute correlation resulted in increased median correlation in controls, but this effect was negligible in the PD group, resulting in progressively larger deviation from the controls with increasing correlation threshold. The loss of correlation between promoter H3K27ac and gene expression in PD replicated in both the PW and NBB cohorts. Importantly, similar results were obtained based on RLE-normalized ChIP-seq data (Supplementary Fig. S8), indicating that the observed decreased correlation between H3K27ac and transcript levels in PD was not driven by our normalization method. These results suggest that, the coupling between promoter H3K27 acetylation state and proximal gene transcription is impaired in the PD prefrontal cortex.

Discussion

Our study addresses the pertinent question of whether aberrant histone acetylation occurs in the brain of individuals with PD. We report hyperacetylation of multiple histone sites in the PD prefrontal cortex with the most pronounced changes observed for H3K27. Our findings, replicated in two independent cohorts, indicate that H3K27 hyperacetylation in the PD brain is a genome-wide phenomenon, but shows a clear predilection for genes previously implicated in PD. Intriguingly, our results also strongly suggest that promoter H3K27 acetylation is decoupled from gene expression in PD.

Our study was motivated by the hypothesis that mitochondrial dysfunction, and in particular respiratory complex I deficiency, may induce a hyperacetylation-prone cellular environment in the PD brain. Exploratory immunoblotting analyses indicated hyperacetylation of multiple histone sites on histones H2B, H3 and H4, with the strongest change observed for H3K27, a marker of active promoters and enhancers with a fundamental role in regulating gene expression[42,46]. Importantly, our experiments indicated that this hyperacetylation was not due to differences in cellular composition of the samples and is unlikely to be induced by anti-Parkinson drugs. The observed upregulation of the histone deacetylases SIRT1 and SIRT3 corroborates previous observations[47] and is compatible with decreased sirtuin activity due to NAD^+ deficiency[22,23,47]. Interestingly, a recent study[51] in rat nigral dopaminergic cells identified H3K27 hyperacetylation as the most prominent histone modification caused by rotenone-induced complex I inhibition, or by knocking-out *TFAM*, encoding mitochondrial transcription factor A, an essential factor for mtDNA maintenance and mitochondrial respiration[50] . While these findings support our hypothesis of mitochondria-dependent dysregulation of histone acetylation in PD, further evidence is needed in order to assess causality.

Our immunoblot findings prompted us to map the genomic landscape of H3K27ac in PD. ChIP-seq analysis corroborated our immunoblot results, showing an increase in the total number of reads in peaks, and pointed out to genome-wide H3K27 hyperacetylation in PD. This is supported by multiple observations: (a) When the H3K27ac peaks were called separately for PD and control samples, the PD group exhibited increased total number of peaks, increased percentage of unique peaks and increased genome coverage. (b) Both cohorts showed higher read counts across most genomic regions in the PD samples. (c) The differential H3K27 acetylation between PD and controls was strongly associated with the effective gene length. This association was positive for hyperacetylated regions and negative for hypoacetylation regions. This is what would be expected in case of general predisposition for hyperacetylation, which by chance is more likely to affect genes occupying larger genomic space.

In addition to the genome-wide increase in H3K27 acetylation, our findings identify 277 genes with altered H3K27 acetylation that replicate independently across the PW and NBB cohorts. Strikingly, 197 of these 277 genes harbor DARs within the same genomic location and six have a well-established implication in PD. Based on their replicability across different populations, and enrichment in PD-implicated genes, we postulate that these regions are specifically targeted in PD.

Meta-analysis combining the cohorts identified a total of 7,551 differentially acetylated peaks with a significant metaP value, mapped to 2,979 genes. We show that these have a strong predilection for genes with an established link to idiopathic and/or familial PD (Table S8), as well as for genes implicated in more complex forms of parkinsonism. Selected genes which have been extensively characterized with respect to PD and neurodegeneration are discussed briefly below.

SNCA encodes α-synuclein, the principal component of Lewy pathology, which is the neuropathological hallmark of PD[2,40]. In addition, *SNCA* mutations cause autosomal dominant PD and genetic variation is associated with idiopathic PD on GWAS. *MAPT* encodes the microtubule associated protein tau, which is the main constituent of neurofibrillary tangles, a typical pathological feature of PD and other parkinsonisms[2,40]. Furthermore, common genetic variation in the *MAPT* gene has been shown to influence the risk of idiopathic PD[50]. *APP* encodes the precursor protein of beta-amyloid (Aβ), the deposition of which is associated with cognitive decline and dementia in PD[41]. Mutations in *PRKN* and *PARK7*, encoding Parkin and DJ-1, respectively, cause autosomal recessive PD and mutations in *FBXO7*, encoding F-box protein, cause juvenile-onset atypical parkinsonism[51]. These three proteins regulate the processes of mitochondrial quality and control and mitophagy, which have been also implicated in idiopathic PD[52,53]. *POLG* encodes DNA polymerase gamma, which replicates and repairs mtDNA. *POLG* mutations cause severe nigrostriatal degeneration with or without clinical parkinsonism[54–56], whereas variation in the gene's polyglutamine tract length has been associated with idiopathic PD[57,58].

Another gene of interest, which is robustly hypoacetylated in our data both at the gene and DAR level, is *PTPRH*, encoding protein tyrosine phosphatase receptor type H. While not much is known about the function of this gene, it was previously associated with PD based on whole exome sequencing analysis[36]. Furthermore, it was shown that RNAi targeting of the *PTPRH* homolog in *Drosophila* enhanced α-synuclein neurotoxicity[36].

The attenuated correlation between H3K27 acetylation and gene expression in our PD samples is intriguing. While the mechanisms underlying this decoupling remain unknown, decoupling between histone acetylation and gene expression was previously observed in cell culture, following widespread histone hyperacetylation, induced by HDAC inhibition[44,45]. Scott et al[44], suggested that this decoupling may be a result of altered miRNA expression,

which in turn downregulates gene expression by transcriptional repression and/or posttranscriptional RNA modulation. While we did not assess miRNA levels in our work, previous studies have demonstrated altered expression of multiple miRNAs in PD[59,60]. Alternatively, it is possible that other other epigenetic mechanisms, such as DNA methylation, play a role in silencing the hyperacetylated PD genome. Further studies are warranted to address the underlying mechanisms and the downstream effects of this decoupling in PD.

In conclusion, we show that H3K27 acetylation, a fundamental mechanism modulating gene expression, is severely dysregulated in the prefrontal cortex of individuals with idiopathic PD. The highly significant concordance between the two cohorts and the clear predilection for genes strongly linked to PD, validate our findings and suggest that epigenetic dysregulation plays an important role in the pathogenesis of the disease. Importantly, since histone acetylation can be modified by drugs modulating the action of histone acetyltransferases and deacetylases, our results open the possibility that agents targeting this modification may have therapeutic potential for PD.

Materials and Methods

Cohorts

Prefrontal cortex (Brodmann area 9) tissue was obtained from two independent cohorts. The discovery cohort (ParkWest, or PW cohort) comprised individuals with idiopathic PD (n = 17) from the ParkWest study, a prospective population-based cohort which has been described in detail[20], and neurologically healthy controls (Ctrl, n = 15) from the Neuromics Biobank for Aging and Neurodegeneration. For a subset of the PW cohort (PD: n =7, Ctrl: n = 7), tissue samples from the striatum (putamen) and cerebellar cortex were also included. All individuals of the PW cohort had undergone whole-exome sequencing[61] and individuals with pathogenic mutations in genes implicated in Mendelian PD and other monogenic neurological disorders had been excluded. All PW individuals were of Western Norwegian origin.

The replication cohort, comprised individuals with idiopathic PD (n = 10) and neurologically healthy controls (n = 11) of Dutch origin, from the Netherlands Brain Bank (NBB, https://www.brainbank.nl/). Demographic information and experimental allocation of all samples is shown in Supplementary Table S1. For each experiment, there were no significant differences for sex, age, or postmortem interval (PMI) between the compared groups.

None of the study participants had disorders or used medication known to influence histone acetylation status, during the last year of life. A complete list of the medication used by the individuals with PD during at least the last 12 months before death can be found in Supplementary Table S1. Individuals with PD fulfilled the National Institute of Neurological Disorders and Stroke[62] and the U.K. PD Society Brain Bank[63] diagnostic criteria for PD at

their final visit. Controls had no clinical or neuropathological evidence of neurodegeneration, or other neurological disease.

Brain tissue

Brains were collected at autopsy and split sagittally along the corpus callosum. One hemisphere was fixed whole in formaldehyde and the other coronally sectioned and snap-frozen in liquid nitrogen. All samples were collected using a standard technique and fixation time of ~2 weeks. There were no significant differences in PMI between patient and control groups in each experiment (Supplementary Table S1). Routine neuropathological examination including immunohistochemistry for α-synuclein, tau and beta-amyloid was performed on all brains. All cases showed neuropathological changes consistent with PD including degeneration of the dopaminergic neurons of the SNc in the presence of Lewy pathology. Controls had no pathological evidence of neurodegeneration.

Chemicals, reagents and media

Unless otherwise indicated, all chemicals were from Sigma Aldrich or from Bio-Rad. Cell culture media and reagents were from Invitrogen.

Cell culture

SH-SY5Y cells (ATCC® CRL-2266™) were cultivated in DMEM /F-12 (1:1) medium with GlutaMAX containing 10 % FBS and penicillin / streptomycin. For differentiation, 8 x 10^5 cells were seeded in a 6-well plate, and at about 80 % confluency sequentially treated with 10 µM retinoid acid in complete medium for 5 days and serum-free medium containing 50 ng / ml BDNF (Tocris) for 3 days as described previously[64]. For pharmacological treatment, fully differentiated cells were exposed to commonly prescribed anti-Parkinson

drugs (APDs) including L-dopa (50 µM), ropinirole (0.05 µM), entacapone (0.05 µM and 0.5 µM) and carbidopa (30 µM and 1.5 µM,) for 48 h, or deacetylase inhibitor trichostatin A (TSA, 0.5 µM) for 8 hours. The applied concentrations were calculated based on the maximum clinical dose according to the Norwegian prescription catalogue (https://www.felleskatalogen.no/medisin). The maximum prescribed dosage was converted to blood concentration, according to reported pharmacokinetic properties in the Norwegian prescription catalogue, and used in the experiment. When the calculated values were below the described IC50 of the compound, an additional sample using the IC50 concentration was included. For L-DOPA, the concentration was chosen based on previous reports[65]. Cell viability was assessed by resazurin assay to optimize drug exposure time. Each cell culture experiment was conducted three times starting with undifferentiated SH-SY5Y cells.

Lysate preparation, protein determination and immunoblotting

Fresh frozen brain tissue was lysed in 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA and 2% SDS. SH-SY5Y cells were lysed in 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 0.2% SDS, 10 µM trichostatin A, 10 µM nicotinamide and complete protease inhibitor cocktail. All samples were homogenized with a 27G syringe and centrifuged at 12.000 x g for 5 min at room temperature. Supernatants were collected and protein concentration determined using the BCA protein kit (Pierce/Thermofisher). Three independently dissected samples of brain tissue (replicates) were analyzed from each individual. Immunoblots were repeated at least 2 times for each marker.

For immunoblotting, 15 µg total protein were separated using AnykD™ MiniProtean® TGX™ Precast Protein gradient gels (Bio-Rad) and transferred to a 0.2 µm PVDF membrane using the Trans-Blot Turbo system (Bio-Rad) at 25 V and 1 Amp for 30 min. After transfer, membranes were blocked with 5% BSA in TBS containing 0.1 % Tween-20 (TBS-Tween) for

antibodies targeting acetylated lysines and in 5% nonfat drymilk in TBS-Tween for all other antibodies. Incubation with primary antibody was carried out overnight at 4 ^{o}C. Primary antibodies were used against acetylated lysine (CST #9441, 1:10.000), histone H2AK5ac (Abcam, ab1764, 1:2000), histone H2BK15ac (CST #9083, 1:10.000), histone H3K9K14ac (CST #9677, 1:20.000), histone H3K27ac (CST #8173S, 1:1000), histone H3K56ac (CST #4243, 1:200), histone H4K5ac (CST #8647, 1:1000), histone H4K12ac (CST #13944S, 1:2000), histone H4K16ac (CST #13534, 1:5000), histone H2A (Abcam, ab18255, 1:1000), histone H2B (Abcam, ab1790, 1:10.000), histone H3 (Abcam, ab1791m 1:20.000), histone H4 (CST # 2935, 1:500), GAPDH (SantaCruz, sc-32233, 1:20.000), NeuN (Millipore, MAB377, 1:1000), GFAP (Sigma-Aldrich, G3893, 1:15.000), CNP1 (Atlas antibodies, AMAb91072, 1:4000), CX3CR1 (Abcam, ab8021, 1:500), NUP188 (Invitrogen, PA5-48940, 1:250), calbindin (Sigma-Aldrich C9848, 1:3000), α-tubulinK40ac (Abcam, ab179484, 1:2000), α-tubulin (Abcam, ab18251, 1:2000), SIRT1 (Sigma Aldrich, S5447, 1:250), SIRT2 (Sigma Aldrich, S8447, 1:2000), SIRT3 (Sigma Aldrich, S4072, 1:2000).

After washing four times with TBS-Tween for 5 min, HRP-conjugated secondary antibodies (rabbit anti-mouse (P0260) or swine anti-rabbit (P0217), DAKO) were incubated for one hour at RT. After washing three times with TBS-Tween and once with TBS for 5 min, immunodetection was carried out using enhanced chemiluminescence (Clarity ECL detection kit, Bio-Rad), images were taken on a ChemiDoc XRS imager (Bio-Rad) and bands were quantified by densitometry using the ImageLab software (Bio-Rad).

To normalize the histone acetylation signal to histone quantity, each acetylation blot was stripped twice using stripping buffer containing 0.2 M glycine, 3.5 mM SDS and 1% Tween-20 at pH 2.2 for one hour at 50 ^{o}C. Complete stripping was confirmed by incubation of the membranes with secondary antibodies only and ECL. No bands were detected even after long exposure time (up to 10 min). After detection, the membrane was washed and re-probed with

primary antibody recognizing the histone being studied for assessment of total histone protein amount. To exclude potential bias due to incomplete stripping, we performed additional experiments where acetylated and total histone markers for each histone modification were assessed in parallel on different blots. To estimate the acetylated fraction of each histone marker, we calculated the signal ratio of acetylated marker to total histone protein. This ratio is referred to as the "acetylation fraction" for each histone marker.

Statistical analysis of Western blot data

Comparison of Western blot densitometric values between PD and controls was performed using linear regression for panacetylation and linear mixed-models for all other markers. Linear mixed-models analysis was chosen in order to account for intraindividual random variability. Subjects were specified as random effects and disease state, sex, age of death, PMI and oligodendrocyte MSP were used as fixed effects. The error distribution was assumed normal, and solution to the mixed models was obtained using the maximum likelihood estimation method. The mixed-models analyses were implemented in SPSS version 22.0.0.

Chromatin Immunoprecipitation sequencing (ChIP-seq)

Frozen brain tissue was shipped to Active Motif where the ChIP-sequencing experiment was performed. For each ChIP-seq experiment approximately 200mg of frozen brain tissue per sample was fixed in PBS + 1% formaldehyde at room temperature for 15 minutes. Fixation was stopped by the addition of glycine to final concentration 0.125 M. The tissue pieces were then treated with a TissueTearer and finally spun down and washed 2x in PBS. Tissue was then homogenized with a lysis buffer using a manual Dounce homogenizer. Lysates were sonicated and the DNA sheared to an average length of 300-500 bp. 14 ug of chromatin was precleared with protein A agarose beads (Invitrogen). Protein-DNA complexes were immuno-

precipitated using 4 ug of H3K27Ac antibody (Active Motif, cat# 39133, lot# 01518010). Complexes were washed, eluted from the beads with SDS buffer, followed by de-crosslinking with proteinase K and RNase treatment overnight at 65 C. The ChIP DNA was purified by phenol-chloroform extraction and ethanol precipitation. Quantitative PCR (qPCR) reactions were carried out in triplicate on specific genomic regions using SYBR Green Supermix (Bio-Rad). The resulting signals were normalized for primer efficiency by carrying out qPCR for each primer pair using Input DNA. Illumina sequencing libraries were prepared from the ChIP and Input DNAs by the standard consecutive enzymatic steps of end-polishing, dA-addition, and adaptor ligation. After a final PCR amplification step, the resulting DNA libraries were quantified and sequenced on Illumina's NextSeq 500 (75 base pair reads, single end).

ChIP-seq quality control and filtering

Raw FASTQ files were trimmed using Trimmomatic version 0.38[66] with options ILLUMINACLIP:TruSeq2-SE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:25. Assessment of FASTQ files was carried out with FastQC version 0.11.8 [http://www.bioinformatics.babraham.ac.uk/projects/fastqc] before and after trimming. Reads were aligned to the GRCh37 human genome using bwa version 0.7.17[67]. Reads with mapping quality < 30 were discareded using samtools version 1.9 [http://www.htslib.org/] and duplicated reads removed using MarkDuplicates from Picard Tools version 2.18.27 [http://broadinstitute.github.io/picard/]. BAM files were converted into BED files and sorted.

Peak calling and read counting

Peak calling procedure was carried out in two complementary approaches: 1) **Per group (PD or controls):** the peaks are called separately for each group, aggregating the relevant samples and inputs. While this approach is not optimal for differential peak calling analysis as it requires harmonization of peaks between the two groups, it can be used to compare the total genome coverage as well as unique H3K27ac genomic regions in each group. 2) **All samples combined:** the peaks are called by aggregating all samples and input controls. This approach does not allow identification of unique genomic regions however, it provides a better identification of consensus peaks, and thus, was used for differential peak analysis.

Specifically, BED files for single samples were merged into the three groups: (a) all PD samples, (b) all control samples (c) all samples. MACS2 version 2.1.1[68] was used on each of the three group BED files without the shifting model (--nomodel) and with the following parameters: sequencing tag size (-s) 75, read extension (--extsize) 200, read shift (--shift) 0, and keeping all duplicates (--keep-dup all). The resulting peak-sets were filtered using the ENCODE black listed regions and only peaks within canonical chromosomes were kept for downstream analyses. Sample-specific enrichment on the identified peaks was performed using featureCounts version 1.6.4[69] with default parameters. To this end, reads within filtered BAM files were extended to reach the fragment length (200) before quantification. For both the individual group and all samples combined approaches, peaks with non-adjusted p-value > 10^{-7} were excluded from the analysis. Reads inside peaks (RiP) were quantified using "featureCounts" program from the "subread" package v2.0.0.

Calculation of Marker Site Profiles (MSP)

In order to identify brain cell-type specific H3K27ac regions we first analyzed H3K27ac ChIP-seq data from $NeuN^+$ and $NeuN^-$ brain cells[25]. Cell-type based broadPeaks (CellType_peak-set), BAM files and metadata files were downloaded from

https://www.synapse.org/#!Synapse:syn4566010. Differential regions between $NeuN^+$ (neurons) and $NeuN^-$ (glia) cells were calculated using "DESeq2" R package, including chromatin amount, library batch, sex, hemisphere, age and pH as covariates in the model. Peaks were defined as cell-type specific differentially acetylated regions (DARs) if they met the following criteria: 1) |fold change| > 4 and 2) mean count > 1000. Peaks were annotated to genes using the "build_annotations" function from "annotatr" R package based on UCSC hg19 genome assembly. Peaks were annotated to all genes for which they intersected a region between 5kb upstream from their transcription start site (TSS) to the end of their 5'UTR. In the next step, we intersected the genes with DARs between glia and neurons with expression-based cortical marker gene lists based on NeuroExpresso database[27]. The DARs were next reassigned to specific glial and neuronal cells if they were annotated to genes defined as cell-type specific based on NeuroExpresso. For example, all DARs annotated to *MBP*, defined as oligodendrocyte marker gene based on NeuroExpresso, were defined as oligodendrocyte marker sites (MSS). In the same way, all DARs annotated to *SLC1A3*, defined as astrocyte marker gene based on NeuroExpresso, were defined as astrocyte MSS. In the next step, the reads from our samples were quantified in regions defined by the CellType_peak-set. The corresponding RiP were then converted to counts per million (CPM) and transformed to log2(CPM + 1). Next, for each cell-type specific MSS, we performed principal component analysis based on the relevant peaks using "prcomp" function from the "stats" R package using (scale = T), as described in[27]. Marker Site Profiles (MSP) were defined as the scores of the samples in the first principal component, transformed to {0-1} range for visualization purposes.

Identification of differentially acetylated regions

In order to further increase the reliability of analyzed peaks, for differential peak analysis, we removed peaks with maximal coverage of < 5 reads/base, calculated as $\max(Peak_{counts})/(Peak_{length})$. Housekeeping normalization factor was calculated as follows: for each peak annotated to a house-keeping gene with mean count > 150 (housekeeping peak), for each sample, we first calculated the ratio between the sample count to the mean count of the peak across all samples ($counts_{sample}/counts_{mean}$). The housekeeping normalization factor was then defined as mean of the ratios for all relevant housekeeping peaks. Housekeeping normalization factors were calculated based on two sets of housekeeping genes: 1) housekeeping genes suggested based on stable transcript expression[24] and 2) housekeeping genes manually selected based on known high and global expression level and very high and broad peak signal covering most of the gene body.

Differential peak analysis was performed using "nbinomWaldTest" function from "DESeq2" R package. Dispersions were calculated by using "estimateDispertions" function from "DESeq2" R package, by setting the fitType parameter to "local". Normalization factors calculated by the "estimateSizeFactors" function were replaced with the housekeeping normalization factor based on our manually selected housekeeping gene set (*ACTB, GAPDH, UBC*). Peaks were annotated to genes as described in the previous section. The output of the analysis was obtained using DESeq2 "results" function, setting the cutoff for adjusted p-value to 0.05, and independentFiltering = T.

Association of GWAS-hits and EGL with observed altered H3K27 acetylation

Effective gene length was defined as sum of length of all regions annotated to the gene (including 1Kb upstream to TSS). We assessed the association between altered H3K27 acetylation defined as -log10(metaP) with being annotated to a PD implicated gene (PDgene), EGL and the total number of common regions aligned to the gene using a linear model. The

total number of common regions in a gene was included as a covariate in the model. Namely, for a common region i, annotated to gene j, the -log(metaP) value was modeled as: $-\log10(metaP)_{ij} \sim PDgene_j + \log10(EGL_j) + \sum CommonRegions_j$. For the analysis was performed in each cohort separately, metaP was replaced with cohort-specific p-value, and the total number of common regions was replaced with the total number of regions (peaks) annotated to the gene in the cohort-specific peak-set.

RNA-seq analysis

Total RNA was extracted from prefrontal cortex tissue homogenate for all samples for which ChIP-seq was performed and sequenced using RiboZero™ Gold (Epidemiology) kit (Illumina, San Diego, CA). Sequencing was conducted at HudsonAlpha Institute for Biotechnology Genomic Services Laboratory. A detailed description of the RNA-seq analysis is provided in the Supplementary Methods and is reported elsewhere[70].

Relative cell abundance was estimated using cell-type specific marker genes, as previously described[27,71]. Since significant group effect was found for oligodendrocytes and microglia estimates in PW (Wilcoxon's $p = 2.6 \times 10^{-4}$) and NBB (Wilcoxon's $p = 0.009$) cohorts respectively, these cell types were included at covariates in the analysis of both cohorts. Count matrix and the code necessary to reproduce these data are available on https://git.app.uib.no/neuromics/cell-composition-rna-pd.

Promoter H3K27ac - expression correlation

ChIP-seq and RNA-seq counts were first adjusted to the relevant covariates (age, sex, PMI, batch, oligodendrocyte estimates and microglia estimates). The adjustment was done by re-fitting the counts using the coefficients calculated by DESeq2 with adjusted model matrix in which categorical covariates were set to 0, and the continuous covariates to the mean of the

original values. Next, for each gene, Pearson's correlation between the adjusted ChIP-seq/RNA-seq seq counts was computed using "cor.test" R function, in each group (controls or PD subjects) separately. Only peaks annotated to promoter regions of the genes were included in the analysis. In cases where multiple promoter peaks were defined for a gene, the correlation was calculated for each promoter separately, and the only the highest correlation was kept. This scenario was either a result of multiple promoters due to existence of several splice variants, or in rare cases, multiple peaks in the same promoter. The choice of the highest correlation was performed independently for each group. Since not all genes are expected to exhibit a good correspondence between promoter H3K27ac state and expression level due to existence of multiple levels of regulation, we compared the distribution of correlations in each group for different thresholds of absolute correlation in either of the groups (e.g., cor > |0.1| means that only genes with absolute correlation > 0.1 in at least one of the groups are included in the analysis). Group comparison was performed using Wilcoxon rank sum test, using "wilcox" R function.

Data and code availability

The R code necessary to reproduce the results presented in the paper is available on https://github.com/ltoker/ChIPseqPD. Raw counts for the ChIP-seq analyses are available in the same repository (https://github.com/ltoker/ChIPseqPD/tree/master/data). Additional R code to reproduce the RNA-seq analyses is available on https://git.app.uib.no/neuromics/cell-composition-rna-pd. The corresponding raw data counts are included in the repository (https://git.app.uib.no/neuromics/cell-composition-rna-pd/tree/master/Data).

Acknowledgments

This work is supported by grants from The Research Council of Norway (ES633272) and Bergen Research Foundation (BFS2017REK05). The cell-type specific H3K27ac ChIP-seq data were generated as part of the PsychENCODE Consortium, supported by: U01MH103392, U01MH103365, U01MH103346, U01MH103340, U01MH103339, R21MH109956, R21MH105881, R21MH105853, R21MH103877, R21MH102791, R01MH111721, R01MH110928, R01MH110927, R01MH110926, R01MH110921, R01MH110920, R01MH110905, R01MH109715, R01MH109677, R01MH105898, R01MH105898, R01MH094714, P50MH106934, U01MH116488, U01MH116487, U01MH116492, U01MH116489, U01MH116438, U01MH116441, U01MH116442, R01MH114911, R01MH114899, R01MH114901, R01MH117293, R01MH117291, R01MH117292 awarded to: Schahram Akbarian (Icahn School of Medicine at Mount Sinai), Gregory Crawford (Duke University), Stella Dracheva (Icahn School of Medicine at Mount Sinai), Peggy Farnham (University of Southern California), Mark Gerstein (Yale University), Daniel Geschwind (University of California, Los Angeles), Fernando Goes (Johns Hopkins University), Thomas M. Hyde (Lieber Institute for Brain Development), Andrew Jaffe (Lieber Institute for Brain Development), James A. Knowles (University of Southern California), Chunyu Liu (SUNY Upstate Medical University), Dalila Pinto (Icahn School of Medicine at Mount Sinai), Panos Roussos (Icahn School of Medicine at Mount Sinai), Stephan Sanders (University of California, San Francisco), Nenad Sestan (Yale University), Pamela Sklar (Icahn School of Medicine at Mount Sinai), Matthew State (University of California, San Francisco), Patrick Sullivan (University of North Carolina), Flora Vaccarino (Yale University), Daniel Weinberger (Lieber Institute for Brain Development), Sherman Weissman (Yale University), Kevin White (University of Chicago), Jeremy Willsey (University of California, San Francisco), and Peter Zandi (Johns Hopkins University). We are grateful to

Prof. Laurence A. Bindoff and Dr. Shreejoy Tripathy for the valuable comments and discussions.

Contributions

L.T. conceived, designed and performed the ChIP-Seq and integrative computational analyses. G.T.T. designed and performed the immunoblots and cell-culture experiments. J.S. contributed to conception and implementation of the ChIP-seq and immunoblot analyses. K.H. contributed to data interpretation and critical revision of the manuscript. G.S.N. performed the preprocessing of the ChIP-seq dataset and contributed to the computational analyses. C.D. conceived, designed, supervised the immunoblots and cell-culture experiments, performed part of the immunoblot experiments, contributed to data analyses and interpretation. O.B.T. and G.A. contributed part of the tissue material, supplementary data, and critical revision of the manuscript. C.T. conceived, designed and directed the study, contributed to data analyses and interpretation and acquired funding for the study. L.T., G.T.T., C.D. and C.T. wrote the manuscript with the active input and participation of J.S. and G.S.N. All authors have read and approved the manuscript.

Ethics declarations

Ethics considerations

Informed consent was available from all individuals. Ethical permission for these studies was obtained from our regional ethics committee (REK 2017/2082, 2010/1700, 131.04)

Competing Interests

The authors report no competing financial interests.

References

1. de Rijk, M. C. *et al.* Prevalence of Parkinson's disease in Europe: A collaborative study of population-based cohorts. Neurologic Diseases in the Elderly Research Group. *Neurology* **54**, S21-3 (2000).
2. Dickson, D. W. Parkinson's disease and parkinsonism: neuropathology. *Cold Spring Harb Perspect Med* **2**, (2012).
3. Klein, C. & Westenberger, A. Genetics of Parkinson's disease. *Cold Spring Harb Perspect Med* **2**, a008888 (2012).
4. Nalls, M. A. *et al.* Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol.* **18**, 1091–1102 (2019).
5. Bender, A. *et al.* High levels of mitochondrial DNA deletions in substantia nigra neurons in aging and Parkinson disease. *Nat Genet* **38**, 515–7 (2006).
6. Dölle, C. *et al.* Defective mitochondrial DNA homeostasis in the substantia nigra in Parkinson disease. *Nat. Commun.* **7**, 13548 (2016).
7. Ryan, B. J., Hoek, S., Fon, E. A. & Wade-Martins, R. Mitochondrial dysfunction and mitophagy in Parkinson's: from familial to sporadic disease. *Trends Biochem. Sci.* **40**, 200–210 (2015).
8. Schapira, A. H. *et al.* Mitochondrial complex I deficiency in Parkinson's disease. *Lancet* **1**, 1269 (1989).
9. Flones, I. H. *et al.* Neuronal complex I deficiency occurs throughout the Parkinson's disease brain, but is not associated with neurodegeneration or mitochondrial DNA damage. *Acta Neuropathol* (2017) doi:10.1007/s00401-017-1794-7.
10. Karamanlidis, G. *et al.* Mitochondrial Complex I Deficiency Increases Protein Acetylation and Accelerates Heart Failure. *Cell Metab.* **18**, 239–250 (2013).

11. Singh, P., Hanson, P. S. & Morris, C. M. Sirtuin-2 Protects Neural Cells from Oxidative Stress and Is Elevated in Neurodegeneration. *Park. Dis*. **2017**, (2017).

12. Park, G. *et al.* Regulation of Histone Acetylation by Autophagy in Parkinson Disease. *J Biol Chem* **291**, 3531–40 (2016).

13. Verdin, E. NAD^+ in aging, metabolism, and neurodegeneration. *Science* **350**, 1208–1213 (2015).

14. Anderson, K. A., Madsen, A. S., Olsen, C. A. & Hirschey, M. D. Metabolic control by sirtuins and other enzymes that sense NAD+, NADH, or their ratio. *Biochim. Biophys. Acta* **1858**, 991–998 (2017).

15. Sauve, A. A. Sirtuin Chemical Mechanisms. *Biochim. Biophys. Acta* **1804**, 1591–1603 (2010).

16. Imai, S. & Guarente, L. It takes two to tango: NAD+ and sirtuins in aging/longevity control. *NPJ Aging Mech. Dis.* **2**, 16017 (2016).

17. Peleg, S., Feller, C., Ladurner, A. G. & Imhof, A. The Metabolic Impact on Histone Acetylation and Transcription in Ageing. *Trends Biochem. Sci.* **41**, 700–711 (2016).

18. Harrison, I. F., Smith, A. D. & Dexter, D. T. Pathological histone acetylation in Parkinson's disease: Neuroprotection and inhibition of microglial activation through SIRT 2 inhibition. *Neurosci. Lett.* **666**, 48–57 (2018).

19. Gebremedhin, K. G. & Rademacher, D. J. Histone H3 acetylation in the postmortem Parkinson's disease primary motor cortex. *Neurosci. Lett.* **627**, 121–125 (2016).

20. Alves, G. *et al.* Incidence of Parkinson's disease in Norway: the Norwegian ParkWest study. *J Neurol Neurosurg Psychiatry* **80**, 851–7 (2009).

21. Balan, I. S., Fiskum, G. & Kristian, T. Visualization and quantification of NAD(H) in brain sections by a novel histo-enzymatic nitrotetrazolium blue staining technique. *Brain Res.* **1316C**, 112 (2010).

22. Braidy, N. *et al.* Age related changes in NAD+ metabolism oxidative stress and Sirt1 activity in wistar rats. *PloS One* **6**, e19194 (2011).

23. Koltai, E. *et al.* Exercise alters SIRT1, SIRT6, NAD and NAMPT levels in skeletal muscle of aged rats. *Mech. Ageing Dev.* **131**, 21–28 (2010).

24. Allhoff, M., Seré, K., F. Pires, J., Zenke, M. & G. Costa, I. Differential peak calling of ChIP-seq signals with replicates with THOR. *Nucleic Acids Res.* **44**, e153–e153 (2016).

25. Girdhar, K. *et al.* Cell-specific histone modification maps in the human frontal lobe link schizophrenia risk to the neuronal epigenome. *Nat. Neurosci.* **21**, 1126–1136 (2018).

26. Oldham, M. C. *et al.* Functional organization of the transcriptome in human brain. *Nat. Neurosci.* **11**, 1271–1282 (2008).

27. Mancarci, B. O. *et al.* Cross-Laboratory Analysis of Brain Cell Type Transcriptomes with Applications to Interpretation of Bulk Tissue Data. *eNeuro* ENEURO.0212-17.2017 (2017) doi:10.1523/ENEURO.0212-17.2017.

28. Titus, A. J., Gallimore, R. M., Salas, L. A. & Christensen, B. C. Cell-type deconvolution from DNA methylation: a review of recent applications. *Hum. Mol. Genet.* **26**, R216–R224 (2017).

29. Sun, W. *et al.* Histone Acetylome-wide Association Study of Autism Spectrum Disorder. *Cell* **167**, 1385-1397.e11 (2016).

30. Marzi, S. J. *et al.* A histone acetylome-wide association study of Alzheimer's disease identifies disease-associated H3K27ac differences in the entorhinal cortex. *Nat. Neurosci.* **21**, 1618 (2018).

31. Andrade-Moraes, C. H. *et al.* Cell number changes in Alzheimer's disease relate to dementia, not to plaques and tangles. *Brain J. Neurol.* **136**, 3738–3752 (2013).

32. von Bartheld, C. S., Bahney, J. & Herculano-Houzel, S. The search for true numbers of neurons and glial cells in the human brain: A review of 150 years of cell counting. *J. Comp. Neurol.* **524**, 3865–3895 (2016).

33. Gómez-Isla, T. *et al.* Profound loss of layer II entorhinal cortex neurons occurs in very mild Alzheimer's disease. *J. Neurosci. Off. J. Soc. Neurosci.* **16**, 4491–4500 (1996).

34. Price, J. L. *et al.* Neuron Number in the Entorhinal Cortex and CA1 in Preclinical Alzheimer Disease. *Arch. Neurol.* **58**, 1395–1402 (2001).

35. Wu, H.-C. *et al.* DLG2, but not TMEM229B, GPNMB, and ITGA8 polymorphism, is associated with Parkinson's disease in a Taiwanese population. *Neurobiol. Aging* **64**, 158.e1-158.e6 (2018).

36. Jansen, I. E. *et al.* Discovery and functional prioritization of Parkinson's disease candidate genes from large-scale whole exome sequencing. *Genome Biol.* **18**, 22 (2017).

37. Foo, J. N. *et al.* Genome-wide association study of Parkinson's disease in East Asians. *Hum. Mol. Genet.* **26**, 226–232 (2017).

38. Schneider, S. A. & Alcalay, R. N. Neuropathology of genetic synucleinopathies with parkinsonism: Review of the literature. *Mov. Disord. Off. J. Mov. Disord. Soc.* **32**, 1504–1523 (2017).

39. Reed, X., Bandrés-Ciga, S., Blauwendraat, C. & Cookson, M. R. The role of monogenic genes in idiopathic Parkinson's disease. *Neurobiol. Dis.* **124**, 230–239 (2019).

40. Dickson, D. W. Neuropathology of Parkinson disease. *Parkinsonism Relat. Disord.* **46 Suppl 1**, S30–S33 (2018).

41. Lim, E. W. *et al.* Amyloid-β and Parkinson's disease. *J. Neurol.* **266**, 2605–2619 (2019).

42. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci.* **107**, 21931–21936 (2010).

43. Karlić, R., Chung, H.-R., Lasserre, J., Vlahoviček, K. & Vingron, M. Histone modification levels are predictive for gene expression. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 2926–2931 (2010).

44. Scott, G. K., Mattie, M. D., Berger, C. E., Benz, S. C. & Benz, C. C. Rapid alteration of microRNA levels by histone deacetylase inhibition. *Cancer Res.* **66**, 1277–1281 (2006).

45. Lopez-Atalaya, J. P., Ito, S., Valor, L. M., Benito, E. & Barco, A. Genomic targets, and histone acetylation and gene expression profiling of neural HDAC inhibition. *Nucleic Acids Res.* **41**, 8072–8084 (2013).

46. Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43–49 (2011).

47. Pallàs, M. *et al.* Modulation of SIRT1 expression in different neurodegenerative models and human pathologies. *Neuroscience* **154**, 1388–1397 (2008).

48. Larsson, N.-G. *et al.* Mitochondrial transcription factor A is necessary for mtDNA maintenance and embryogenesis in mice. *Nat. Genet.* **18**, 231–236 (1998).

49. Huang, M. *et al.* Mitochondrial Dysfunction Induces Epigenetic Dysregulation by H3K27 Hyperacetylation to Perturb Active Enhancers in Parkinson's Disease Models. *bioRxiv* 808246 (2019) doi:10.1101/808246.

50. Elbaz, A. *et al.* Independent and joint effects of the MAPT and SNCA genes in Parkinson disease. *Ann. Neurol.* **69**, 778–792 (2011).

51. Shojaee, S. *et al.* Genome-wide linkage analysis of a Parkinsonian-pyramidal syndrome pedigree by 500 K SNP arrays. *Am. J. Hum. Genet.* **82**, 1375–1384 (2008).

52. Pickrell, A. M. & Youle, R. J. The roles of PINK1, parkin, and mitochondrial fidelity in Parkinson's disease. *Neuron* **85**, 257–273 (2015).

53. Burchell, V. S. *et al.* The Parkinson's disease-linked proteins Fbxo7 and Parkin interact to mediate mitophagy. *Nat. Neurosci.* **16**, 1257–1265 (2013).

54. Tzoulis, C. *et al.* Severe nigrostriatal degeneration without clinical parkinsonism in patients with polymerase gamma mutations. *Brain* **136**, 2393–404 (2013).

55. Palin, E. J., Paetau, A. & Suomalainen, A. Mesencephalic complex I deficiency does not correlate with parkinsonism in mitochondrial DNA maintenance disorders. *Brain* **136**, 2379–92 (2013).

56. Luoma, P. *et al.* Parkinsonism, premature menopause, and mitochondrial DNA polymerase gamma mutations: clinical and molecular genetic study. *Lancet* **364**, 875–82 (2004).

57. Luoma, P. T. *et al.* Mitochondrial DNA polymerase gamma variants in idiopathic sporadic Parkinson disease. *Neurology* **69**, 1152–9 (2007).

58. Balafkan, N. *et al.* Number of CAG repeats in POLG1 and its association with Parkinson disease in the Norwegian population. *Mitochondrion* **12**, 640–3 (2012).

59. Leggio, L. *et al.* microRNAs in Parkinson's Disease: From Pathogenesis to Novel Diagnostic and Therapeutic Approaches. *Int. J. Mol. Sci.* **18**, (2017).

60. Uwatoko, H. *et al.* Identification of plasma microRNA expression changes in multiple system atrophy and Parkinson's disease. *Mol. Brain* **12**, 49 (2019).

61. Gaare, J. J. *et al.* Rare genetic variation in mitochondrial pathways influences the risk for Parkinson's disease: Mitochondrial Pathways In PD. *Mov. Disord.* **33**, 1591–1600 (2018).

62. Gelb, D. J., Oliver, E. & Gilman, S. Diagnostic criteria for Parkinson disease. *Arch. Neurol.* **56**, 33–39 (1999).

63. Ward, C. D. & Gibb, W. R. Research diagnostic criteria for Parkinson's disease. *Adv Neurol* **53**, 245–9 (1990).

64. Encinas, M. *et al.* Sequential treatment of SH-SY5Y cells with retinoic acid and brain-derived neurotrophic factor gives rise to fully differentiated, neurotrophic factor-dependent, human neuron-like cells. *J. Neurochem.* **75**, 991–1003 (2000).

65. Ingman, K. *et al.* The effect of different dosing regimens of levodopa/carbidopa/entacapone on plasma levodopa concentrations. *Eur. J. Clin. Pharmacol.* **68**, 281–289 (2012).

66. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

67. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–95 (2010).

68. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

69. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinforma. Oxf. Engl.* **30**, 923–930 (2014).

70. Nido, G. S. *et al.* Common gene expression signatures in Parkinson's disease are driven by changes in cell composition. *bioRxiv* 778910 (2019) doi:10.1101/778910.

71. Toker, L., Mancarci, B. O., Tripathy, S. & Pavlidis, P. Transcriptomic Evidence for Alterations in Astrocytes and Parvalbumin Interneurons in Subjects With Bipolar Disorder and Schizophrenia. *Biol. Psychiatry* **84**, 787–796 (2018).

Tables & Figures

Table 1. Top replicated genes between PW and NBB cohorts

GeneSymbol	log2(FC)_PW	padj_PW	log2(FC)_NBB	padj_NBB
Hypoacetylated genes				
PTPRH	-0.81	9.1E-03	-0.94	0.03
JUP	-0.43	0.05	-0.44	0.02
Hyperacetylated genes				
DLG2	0.68	9.8E-04	0.77	2.0E-03
TNRC6B	0.74	9.8E-04	0.9	0.03
SIK3	0.66	3.2E-03	0.52	0.01
SGK1	0.65	3.2E-03	0.72	0.01
PPP2R2B	1.3	3.5E-03	0.7	8.0E-03
FCHSD2	0.83	3.5E-03	1	0.01
CTNND2	0.52	3.5E-03	0.63	0.01
PTK2	0.71	3.5E-03	0.55	0.02
SLCO1A2	0.58	3.5E-03	0.68	0.02
LPAR1	0.99	3.5E-03	0.55	0.04
FRMD4B	0.64	3.5E-03	0.53	0.04
RNF216	0.6	5.1E-03	0.59	2.3E-03
TMEM132B	0.88	5.1E-03	0.59	0.02
PIP4K2A	0.57	5.1E-03	0.52	0.03
MAP7	0.55	5.9E-03	0.52	0.04
PDE4B	0.59	6.5E-03	0.67	2.4E-03
FRMD5	0.55	6.5E-03	0.67	0.01

Table 2. Genes annotated to the top replicated DARs.

Symbol	Region Annotation	CHR	Region Location	Peak Location (PW, NBB)	MetaP	LFC PW, NBB	PeakName PW, NBB	Proportion Significant
Common hypoacetylated regions								
PTPRH	introns	19	55718568-55720750	55720288-55722414, 55717912-55719103	1.7E-07	-0.81, -0.94	Peak_20188, Peak_15767	3^{p}/4(0.8)
JUP	introns	17	39681526-39923630	39805764-39808133, 39807092-39808006	1.1E-05	-0.43, -0.99	Peak_53983, Peak_81536	$5^{p}/10^{p}$ (0.5)
Common hyperacetylated regions (top 15)								
*DLG2**	exons	11	83874504-83874554	83873551-83879844, 83873666-83877544	7.7E-12	0.68, 1.05	Peak_21808, Peak_40741	$30^{p}/49^{p}$(0.6)
KIT	exons	4	55604595-55606881	55606747-55609274, 55606782-55609553	8.1E-11	0.63, 1.26	Peak_21777, Peak_37103	$2/3^{p}$ (0.7)
CNTNAP2	exons	7	147336198-147336398	147333973-147338260, 147333944-147337213	6.2E-09	0.51, 1.07	Peak_44999, Peak_47588	$11/18^{p}$ (0.6)
RNF216	introns	7	5681008-5692043	5679685-5685127, 5688370-5705729	1.2E-08	0.6, 0.53	Peak_38631, Peak_12618	11^{p}/16 (0.7)
FRMD5	introns	15	44216508-44393656	44260324-44263821, 44244249-44247825	1.6E-08	0.68, 0.67	Peak_18018, Peak_18744	$12^{p}/17^{p}$(0.7)
SOX2-OT	introns	3	181328718-181417385	181358880-181360150, 181341154-181348014	1.7E-08	0.55, 0.91	Peak_43667, Peak_31549	9/12(0.8)
MAGI2	1to5kb	7	78120338-78124337	78121257-78128843, 78115190-78121170	2.2E-08	0.4, 0.93	Peak_18816, Peak_24042	9^{p}/23(0.4)
PPP2R2B	introns	5	146080706-146460621	146383635-146384203, 146459075-146469815	2.8E-08	1.3, 0.52	Peak_135972, Peak_17645	$17^{p}/27^{p}$(0.6)
KCNH1	introns	1	210948887-210970849	210966438-210969215, 210946491-210950181	3.8E-08	0.55, 1.0	Peak_51838, Peak_38212	$9/18^{p}$ (0.5)
CDH1	introns	16	68835797-68842326	68840159-68841130, 68782065-68837136	4.8E-08	0.84, 0.37	Peak_66179, Peak_3389	$5/7^{p}$ (0.7)
*MAP4K4**	introns	2	102315001-102407181	102330963-102334475, 102367013-102375949	5.2E-08	0.54, 0.62	Peak_30093, Peak_9009	$10^{p}/24^{p}$(0.4)
METTL24	introns	6	110644076-110729536	110656829-110661962, 110685537-110688887	5.4E-08	0.48, 0.97	Peak_22739, Peak_27591	$8/13^{p}$(0.6)
PIP4K2A	introns	10	22898647-23003111	22934570-22952399, 22957818-22962517	5.6E-08	0.58, 0.52	Peak_22480, Peak_27954	$9^{p}/18^{p}$(0.5)
SIK3	introns	11	116827781-116968858	116825819-116835556, 116846923-116884577	5.8E-08	0.6, 0.49	Peak_21504, Peak_11226	$13^{p}/16^{p}$(0.8)
DOCK10	introns	2	225796386-225906968	225798681-225800264, 225845772-225849871	8.1E-08	0.61, 0.77	Peak_94440, Peak_22866	$6/10^{p}$(0.6)
Common hyperacetylated regions (additional GWAS hits)								
*CRHR1**	introns	17	43707525-43750192	43726544-43728643, 43741886-43747152	2.0E-07	0.43, 0.83	Peak_74390, Peak_30944	$6/14^{p}$(0.4)
*MBNL2**	introns	13	97873814-97927885	97919075-97922481, 97906667-97916758	1.2E-05	0.44, 0.44	Peak_29656, Peak_16279	$14^{p}/22^{p}$(0.6)

For hyperacetylated regions, the top 15 genes with DARs common to the two cohorts and common DARs in genes previously associated with PD are shown. The genes were ranked based on Fisher's meta p-value. The full list of genes is provided in Supplementary Table S6.
* - Previous association with PD based on GWAS; Region Location – genomic location (on the indicated chromosome) of the annotated region; Peak location – genomic location of the relevant peak from each cohort (upper – PW, lower - NBB); LFC – log of change compared to controls; Peak Name – the peak identifier in each cohort; Proportion significance – DARs/total common regions annotated to the gene; ^{p}- peaks annotated to promoter region of the gene.

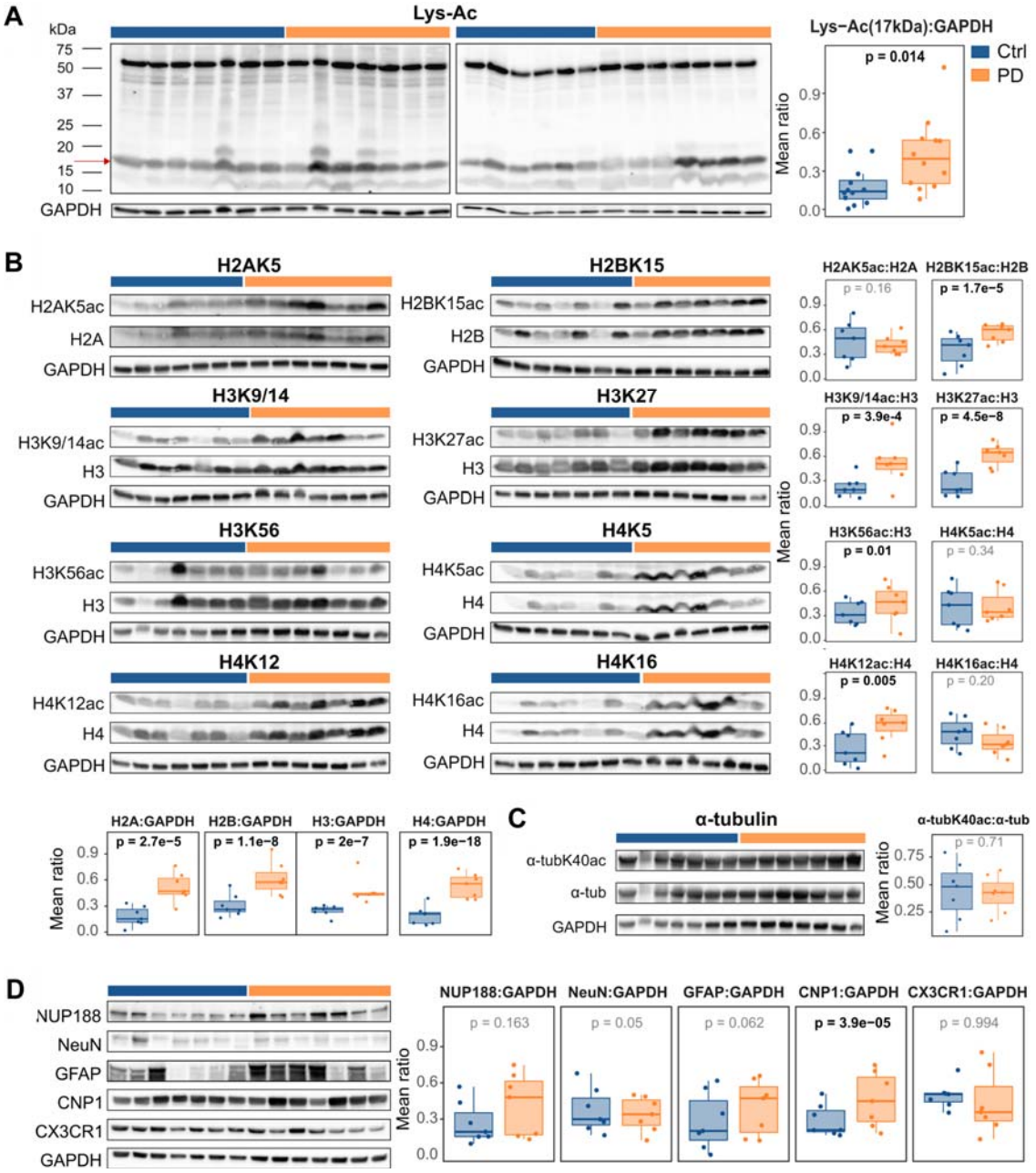

Fig. 1: Hyperacetylation of multiple histone sites in the PD prefrontal cortex

Representative Western blots and the ratios of their quantified densitometric measurements of the indicated proteins (or their modifications) in individuals with PD (orange) and controls (blue) are shown. GAPDH served as loading control. Values were rescaled 0-1. Boxplots show the mean of replicates for each individual. The indicated p-values are based on a linear mixed model (see methods). **A.** Global protein acetylation detected by a pan-acetyl lysine antibody. **B.** Relative acetylation of histone lysine residues and quantification of the total histone proteins. **C.** Relative acetylation of α-tubulin on lysine K40. **D,** Immunodetection and densitometric quantification of relative protein levels of the nuclear marker (NUP188) and cell type-specific marker proteins: NeuN – neurons, GFAP – astrocytes, CNP1 – oligodendrocytes, CX3CR1 - microglia.

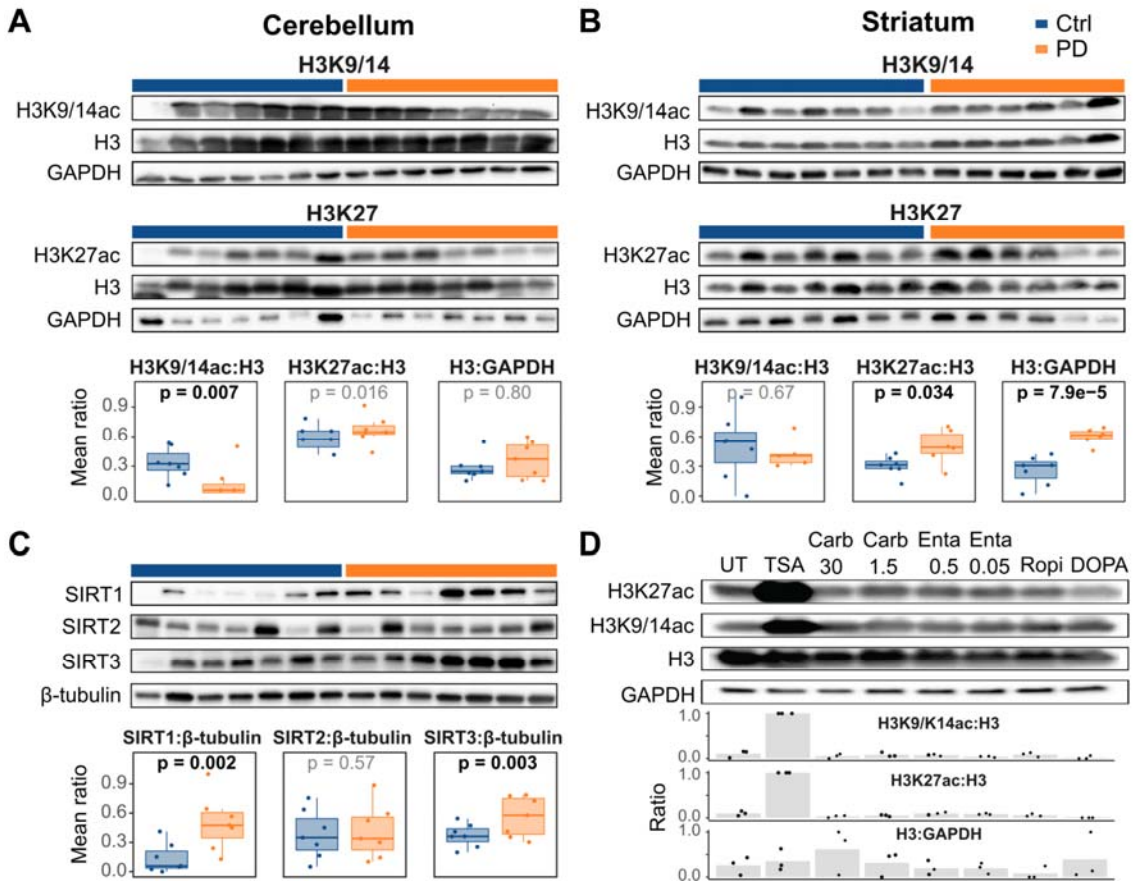

Fig. 2: Altered H3 acetylation occurs in multiple brain regions, is accompanied by sirtuin upregulation and is not induced by anti-Parkinson drugs in-vitro

A-C, Representative Western blots and quantification plots showing the ratios of quantified densitometric measurements of the indicated proteins in individuals with PD (orange) and controls (blue). GAPDH and β-tubulin serve as loading controls. Values were rescaled 0-1. Boxplots show the mean of replicates for each individual. The indicated p-values are based on a linear mixed model (see methods). **A, B,** Acetylated H3K9/14 and H3K27 and total histone protein H3 levels in the cerebellum (A) and striatum (B) from individuals with PD and controls. **C.** Sirtuin proteins in the prefrontal cortex of PD and controls. **D,** Acetylated H3K9/14, H3K27 and total H3 protein in differentiated SH-SY5Y-derived dopaminergic neuron-like cells treated with the histone deacetylase inhibitor Trichostatin A (TSA) or anti-Parkinson drugs used by the PW patients during their last year of life. The numbers indicate the concentration (in µM) of the drug. Carb: Carbidopa; Enta: Entacapone; Ropi: Ropinirole; DOPA: L-Dopa. UT: untreated control. A representative blot is shown. Bar plots show the mean quantified ratios of three biological replicates (rescaled 0-1). Dots indicate the values of each of the individual replicates.

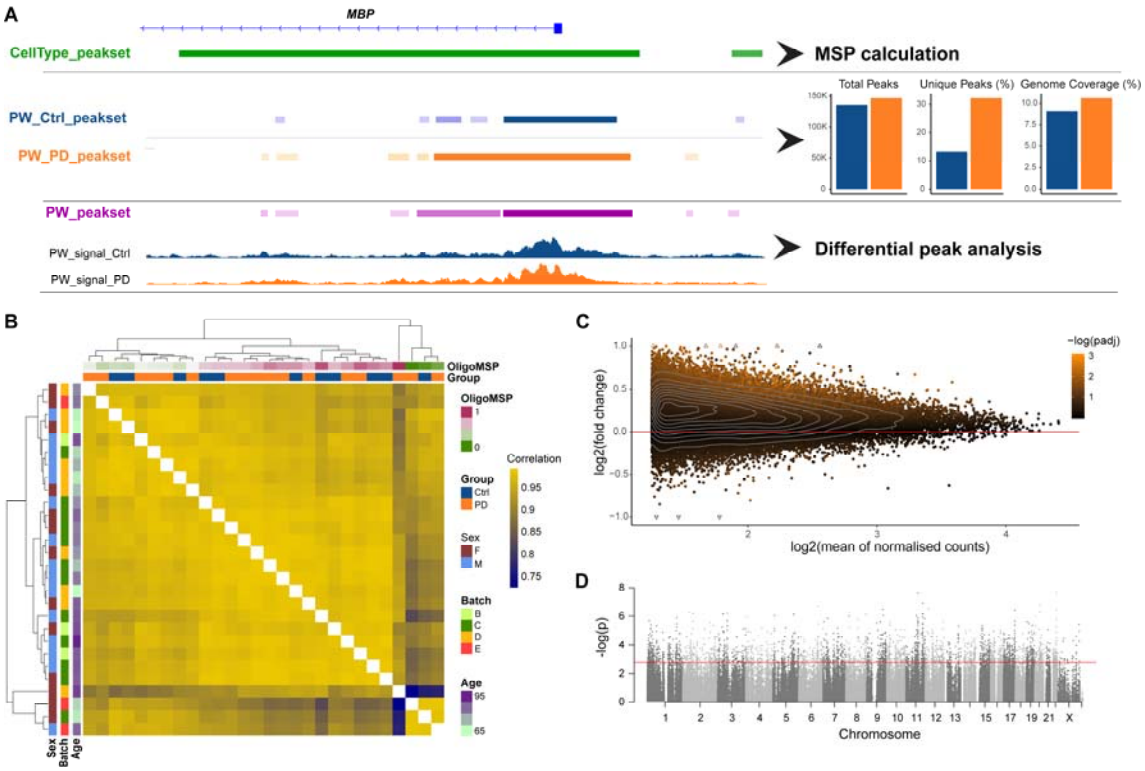

Fig. 3: H3K27ac ChIP-seq analysis of PD and control samples.

A. Different peak-sets used for the different stages of ChIP-seq analysis. Shown, as an example, is a genomic region around the oligodendrocyte marker *MBP*. Peaks annotated to this region, in each peak-set are shown by colored boxes. The CellType_peak-set was used for the MSP calculation. Group specific peak-sets were used to characterize the distribution of H3K27ac bound regions in cases and controls. Barplots show the comparison between the two groups (Supplementary Table S3). PD group exhibited higher number of peaks, higher percentage of unique peaks and higher percentage of genomic coverage, consistent with genome-wide H3K27 hyperacetylation. The PW_peak-set was used for the differential peak analysis. Shown are group overlays of sample fold of change compared to input control. **B.** Hierarchical clustering of the samples indicated that samples cluster based on their cellular composition. Supplementary Fig. S6 shows the association of each of the variables with the main principal components of the data. **C.** MA plot based on the PW_peak-set indicates that increased H3K27ac in PD is observed genome-wide, rather than being restricted to specific regions. **D,** Manhattan plot showing the distribution of genomic locations and differential p-values of the H3K27ac PW_peak-set. The red dashed line indicates the -log10 of the highest p-value with < 5% false positive rate (Benjamini-Hochberg adjusted p-value < 0.05). Corresponding plots for the NBB cohort are shown in Supplementary Fig. S5.

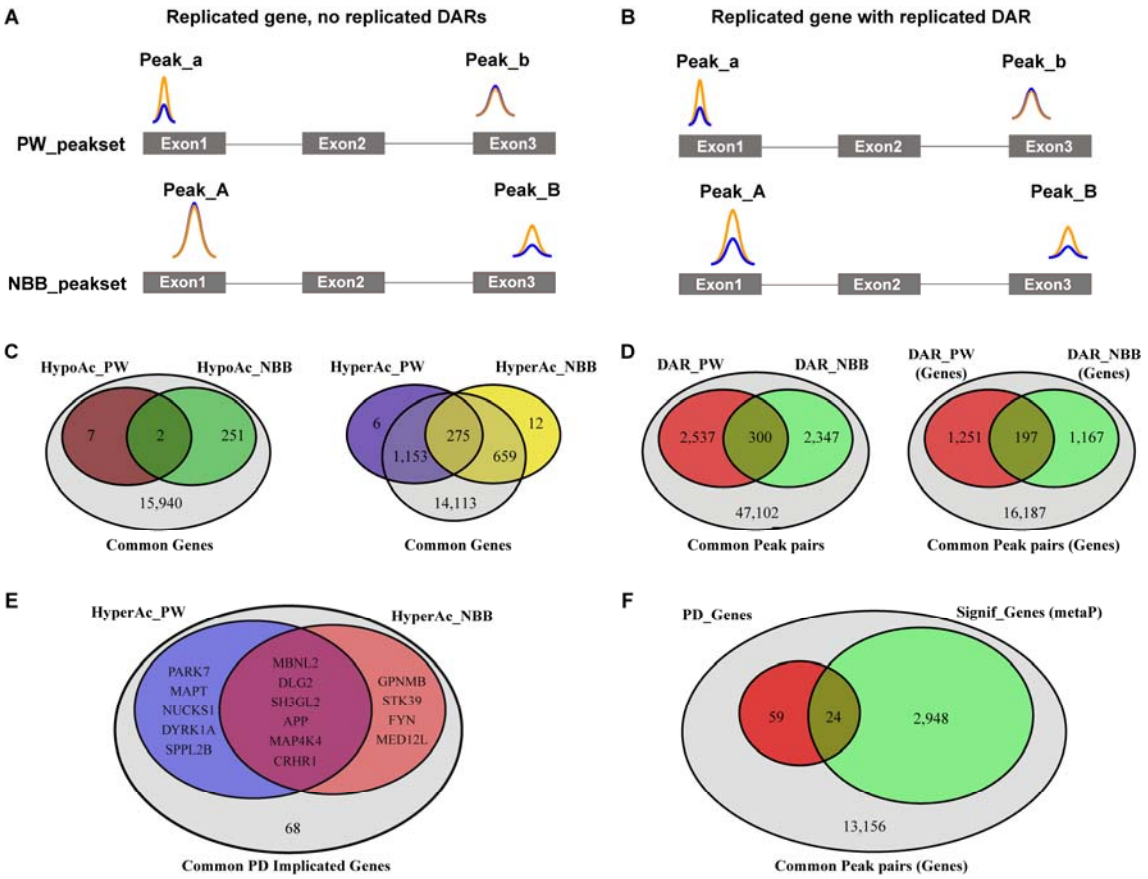

Fig. 4: Replicated genes and replicated DARs

A,B. Schematic representation of two levels of replication between the PW and NBB cohorts. H3K27ac peaks are illustrated by colored lines, representing control (blue) and PD (orange) samples, with the height of the peak representing the peak intensity. Shown is a hypothetical gene X, with exons represented as grey boxes and introns represented by lines. In both cases, the gene has two pairs of common peaks: Peak_a/A, mapped to the first exon of the gene and Peak_b/B, mapped to the third exon of the gene. **A.** Altered H3K27 acetylation occurs at different regions of gene X in each cohort. Thus, while gene X harbors altered H3K27 acetylation in both cohorts (i.e. gene X is a replicated gene), it does not harbor replicated DARs. **B.** Altered H3K27 acetylation occurs at the same region of gene X in both cohorts. In this case, gene X is a replicated gene with a replicated DAR. **C.** Venn diagram showing the number of replicated genes (with or without replicated DARs) between the two cohorts, out of all genes represented in our ChIP-seq data from both cohorts. **D.** Venn diagrams showing the number of replicated DARs between the two cohorts, out of all common peak pairs (left) and their corresponding annotated genes (right). **E.** Venn diagram showing the overlap of PD-implicated genes with DARs between cohorts. All DARs exhibited hyperacetylation. **F.** Venn diagram showing the number of PD-implicated genes with adjusted metaP < 0.05.

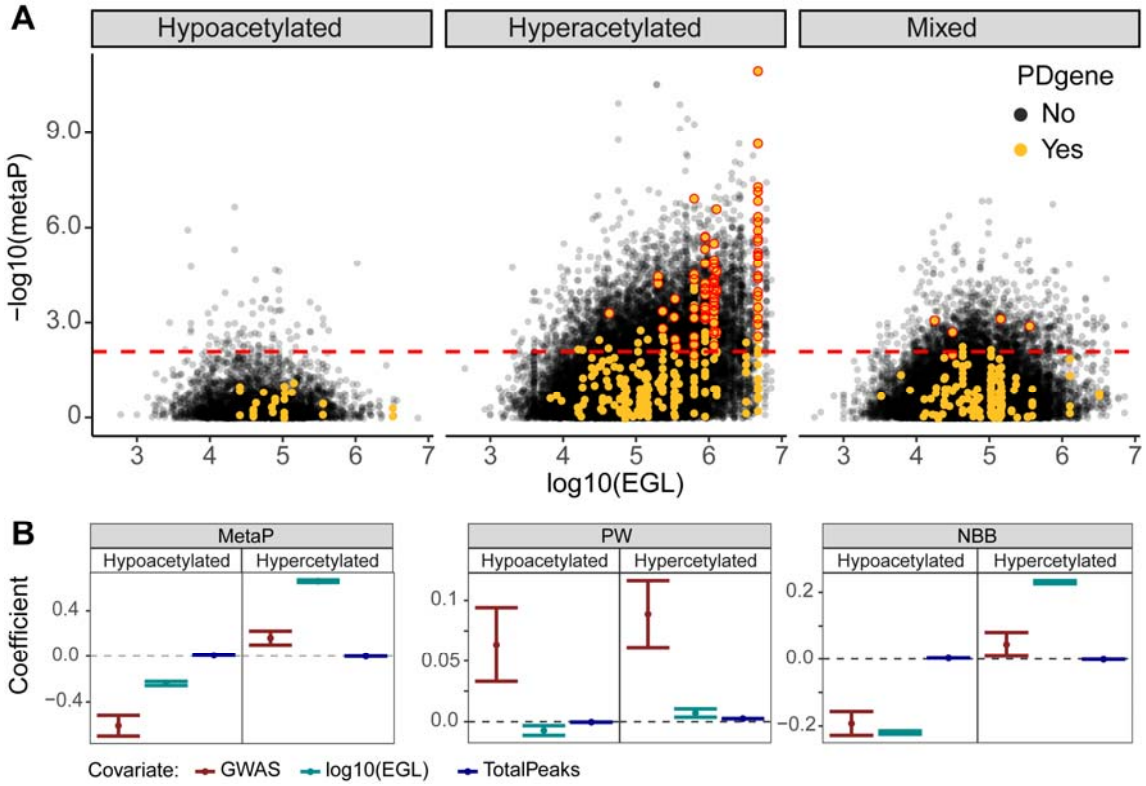

Fig. 5: The H3K27 acetylation state is strongly associated with PD-GWAS hits and the effective gene length

A. Common regions were divided into three groups: Hypoacetylated – regions showing hypoacetylation in both cohorts; Hyperacetylated – regions showing hyperacetylation in both cohorts; Mixed – regions with opposite acetylation trends in the two cohorts. The dashed line represents the -log10 of the highest metaP corresponding to adjusted metaP < 0.05. Multiple common regions within the same gene are shown. Highlighted are regions identified as DAR in at least one of the cohorts within PD-GWAS hits. metaP – Fisher's meta-p value of the corresponding two peaks from PW and NBB cohorts. **B.** Association between -log(p) and each of the indicated covariates based on linear model. Shown are the coefficients and 95% confidence intervals. The results are shown for the common regions (metaP) and each of the cohorts separately. Peaks annotated to miRNAs and snRNAs were excluded from the analysis. GWAS – a gene is a PD-GWAS hit. EGL – effective gene length. TotalPeaks – total number of common regions or peaks annotated to the gene.

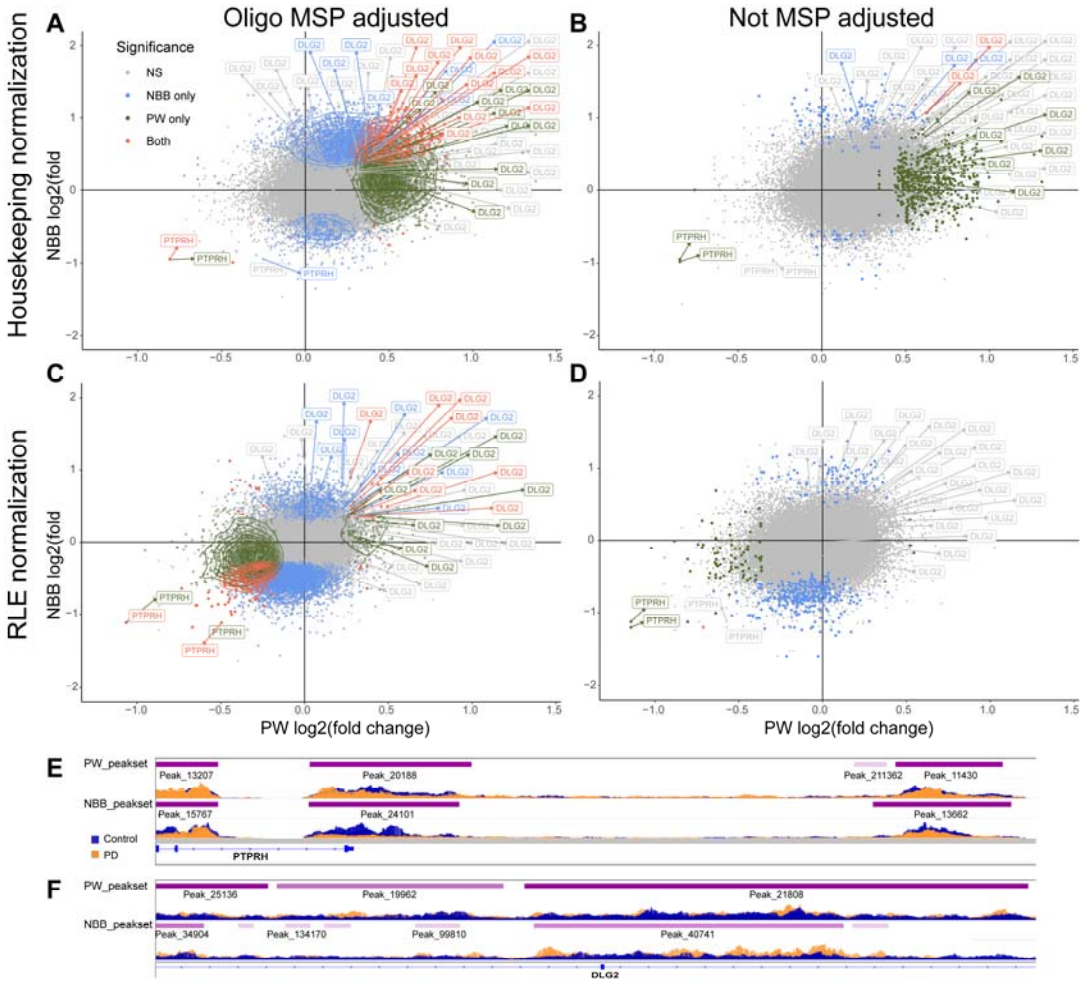

Fig. 6: Effect of normalization method and cell composition correction on H3K27ac ChIP-seq outcome

A-D. Log2 fold change of comparable H3K27ac regions from NBB (y axis) and PW (x axis) were plotted against each other. The color indicates whether and in which of the cohorts the region was identified as DAR. Both cell-type correction (**A,C** vs**. B,D**) and normalization method (**A,B** vs. **C,D**) have a dramatic effect on the outcome of the analysis. **E, F** – IGV tracks for two of the common DARs are shown for *PTPRH* (hypoacetylated, **E**) and *DLG2* (hyperacetylated, **F**) genes. Y-axis represents the overlay of fold of change relative to input control.

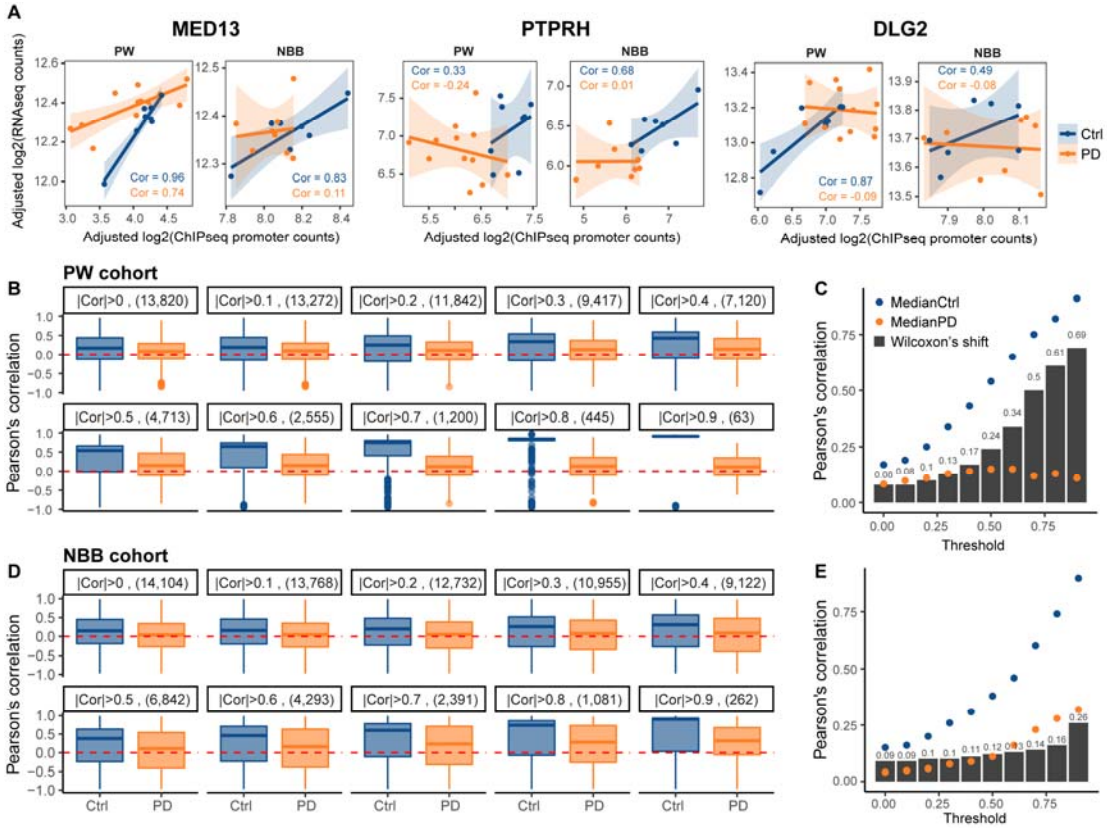

Fig. 7: Decreased correlation between promoter H3K27 acetylation state and gene expression in PD

Pearson's correlation between the adjusted promoter H3K27ac counts and the adjusted RNA-seq counts was calculated for each gene across control or PD subjects in each cohort separately. **A.** Representative scatter plots showing the adjusted promoter H3K27ac counts (X-axis), the adjusted RNA-seq counts (Y-axis) and the calculated Pearson's correlation for three genes: *MED13*, *PTPRH* and *DLG2* in control (blue) or PD (orange) individuals from the PW or NBB cohorts. Each dot represents an individual. Cor - Pearson's correlation. **B-E,** The distribution of correlations for all compatible genes was compared between the groups for various thresholds of minimal absolute correlation in either of the groups. Results are shown for PW (**B,C**) and NBB (**D,E**) cohorts. The number of genes that remained after applying the correlation threshold is shown in parentheses. The correlations remained closely distributed around 0 in PD subjects from both cohorts, regardless of the correlation threshold. **C,E,** Median correlations in each group and the calculated Wilcoxon's delta shift for the different correlation thresholds.